基于深度神经网络的图像缺损修复方法综述

李月龙 高 云 闫家良 邹佰翰 汪剑鸣

(天津工业大学天津市自主智能技术与系统重点实验室 天津 300387)
(天津工业大学计算机科学与技术学院 天津 300387)

摘 要 图像缺损修复研究旨在通过计算机自动修复图像中的缺损内容。近年来，深度神经网络技术的出现有效促进了相关研究的发展。本文针对该类研究进行了系统梳理和综合介绍。依据网络架构类型，具体将方法分为五类：Context-Encoder 类、U-Net 类、CGAN 类、DCGAN 类以及 StackGAN 类。我们具体分析了每类方法的思路、特点、优势和缺陷，并基于系统性实验，在公开大规模数据集上客观对比评价每一类方法的精度和性能。最后对目前相关工作中存在的问题和挑战进行了阐述和介绍。

关键词 图像修复；图像缺损；深度神经网络；计算机视觉；图像处理
中图法分类号 TP391 **DOI 号** 10.11897/SP.J.1016.2021.02295

Image Inpainting Methods Based on Deep Neural Networks: A Review

LI Yue-Long GAO Yun YAN Jia-Liang ZOU Bai-Han WANG Jian-Ming

(*Tianjin Key Laboratory Autonomous Intelligent Technology and Systems, Tiangong University, Tianjin* 300387)
(*School of Computer Science and Technology, Tiangong University, Tianjin* 300387)

Abstract At present, the image is one of the most important tools of visual information storage and transmission. Image inpainting is a research that aims to autonomously repair defected and impaired image content by computer. This is an interesting research topic attracting the attention of computer vision, image processing, and machine learning field. Since the content of impaired area is hardly predicted beforehand, a smart solution method must have the ability to model complex general image information. However, complex information modeling is always a tough mission in visual researches. Thus, related researches once progressed quite slow in decades of years. But things are different because of the development of deep neural networks, which is well known for the ability of complex information learning and modeling. Nowadays, deep neural network based approaches have become the overwhelming solution of image inpainting tasks, and really advanced the general technical merit. In this paper, state of the art works of deep neural network based image inpainting is systematically combed and comprehensively introduced. Since the main difference among those related approaches comes from the network structure, we will class and discuss them according to network architecture discrepancy. Specifically, all mentioned methods are divided into five categories, namely, Context-Encoder class, U-Net class, CGAN class, DCGAN class, and StackGAN class. The Context-Encoder is a famous context information modeling network structure, which can effectively model general visual content. The U-Net is a network construction that can fuse multiple

收稿日期：2020-02-13，在线发布日期：2020-09-13.本课题得到国家自然科学基金（61771340）、天津市自然科学基金（18JCYBJC15300、19JCYBJC15600）、天津市高等学校创新团队培养计划（TD13-5032）、天津市科技计划项目（19PTZWHZ00020）资助.**李月龙**，博士，教授，博士生导师，天津市特聘教授青年学者，中国计算机学会（CCF）会员，主要研究方向为计算机视觉、机器学习、模式识别、图像修复、遮挡重建、轮廓提取、人脸识别. E-mail: liyuelong@pku.edu.cn. **高 云**，硕士研究生，主要研究领域为机器视觉、图像处理. **闫家良**，硕士研究生，主要研究方向为计算机视觉、模式识别. **邹佰翰**，本科生，主要研究方向为目标识别、图像分割. **汪剑鸣**（**通信作者**），博士，教授，计算机学会（CCF）会员，主要研究领域为计算机视觉. E-mail: wangjianming@tiangong.edu.cn.

scale level knowledge, which contains many down and up sampling. The CGAN is a conditional version of traditional GAN structure. This strategy promotes the pertinence and controllability of GAN, which is famous for the brand new adversarial mechanism and powerful synthesizing capability. The DCGAN is a deep level fusion of classical CNN and GAN, and it is a reinforced instrument of feature extraction and modeling. Those structures have evident composition characteristics and are fundamentally distinct from each other. But all of them have been successfully utilized to solve the puzzle of image inpainting, and great outcomes have been achieved. And we believe they stand for the highest technique level of current research to image inpainting problems. Hence, in this paper, we will be detailed discuss the ideas, features, advantages, and defects of each of the five categories, and they will be directly compared and assessed. In order to fairly and objectively measure and evaluate the mentioned methods, carefully designed experiments are conducted on publicly available large-scale databases. Both quantitative and qualitative performance will be demonstrated and discussed to analyze the approaches comprehensively. Finally, the remaining problems and challenges of current works are elaborated and discussed.

Keywords Image inpainting; Impaired images; Deep neural network; Computer vision; Image processing

1 引 言

图像是人类最重要的信息表达和传递方式之一，对人类的沟通交流有着不可替代的突出作用. 然而，在图像信息传递和存储过程中，不可避免地会出现损失，影响信息呈现质量. 在目前使用最为普遍的数字图像（相对于逻辑图像而言）中，信息缺损最主要的表现形式就是像素丢失，如图 1 所示.

从图中可以明显看出，信息缺损对图像内容的呈现产生严重影响. 这是由于图像属于非结构性信息，其中各部分内容的相对独立性较差，所以即使是尺寸很小的缺损（如第 1 列中，图像仅缺损 10% 内容），也会给人图像整体质量较差的直观感受. 显而易见，存在内容缺损的图像是不能直接呈现给读者和用户的. 在图像采集设备不断普及的今天，各类图像数量每天成几何级数级增长，研究设计能够智能修复图像缺损内容的自动化方法和技术势在必行.

由于缺失的信息不可能无中生有，应对图像内容缺损，最关键的核心工作就是如何获取补偿信息. 众所周知，图像可以表达的内容包罗万象，颜色、纹理、结构、组成等各方面均蕴含大量可变性，因此对真实图像的补偿知识建模属于对超大规模信息的高层次复杂结构抽象化凝练问题. 在有限存储空间和运算能力前提下，开展该类工作极为困难. 所以在过去相当长一段时间内，图像缺损自动修复问题曾发展非常缓慢，直至深度神经网络技术的出现.

深度神经网络[1]以结构庞大变量众多闻名，客观上具备存储固化大规模信息知识的可能，而且其个性化信息响应方式使得其可以依据不同图像灵活运用补偿知识. 因此近年来基于深度神经网络实现图像缺损内容修复成为该领域最主要的新兴研究方向，优秀成果不断涌现[2-37]，有效促进了该领域整体研究水平的快速提升.

图 1 图像缺损示例（第 1 行为中心区域缺损，由左至右缺损比例分别为 10%，15%，20%，25%；第 2 行为随机区域缺损，由左至右缺损比例分别为 10%，20%，30%，40%）

然而，基于前期大量相关调研，我们发现目前专门针对基于深度神经网络缺损图像修复技术的综述性工作偏少. 现在能够查询到的公开发表工作仅有两个：Elharrouss 等[38]对图像缺损修复问题进行了已有解决思路的全面综述，但缺少对基于深度神经网络技术的专门化系统分析总结和对比归纳；而在另一文献中，强振平等[39]虽然针对深度神经网络技术在图像缺损问题上的修复应用进行了系统性研究探索，但遗憾的是没有提供任何实验表现支撑，对方法的分析深度有限. 显然，该现状与深度神经网络技术在该领域内的垄断性优势地位并不匹配.

基于此，本文拟针对基于深度神经网络的图像缺损自动修复方法开展专门系统性综述研究. 从方法构成的本质性特征出发，我们选择依据网络核心架构的不同对方法进行分类归纳总结和分析，并依托在公开大规模数据集上的系统性对比实验进行精度和性能的客观评价. 其中包含一些我们的个人观点和看法，未必完全正确，但希望能够对有意了解该研究方向的读者起到微薄帮助.

本文的主要贡献在于：

（1）对目前解决图像缺损修复问题最有效的 state of the art 方法，即深度神经网络类方法，进行了系统性分析和归纳总结；

（2）对当前评价图像缺损修复方法的公开数据集、精度和性能评价指标等进行系统性归纳介绍；

（3）在大规模公开数据集上开展分析评价实验，客观度量方法性能，并进行综合性能对比，探究方法成功和失败背后的核心原因；

（4）系统性分析列举了当前研究所面临的挑战性问题和核心技术难点.

本文余下章节的内容组织安排如下：第 2 节介绍相关基础知识；第 3 节对深度神经网络类图像缺损修复方法进行系统性归纳和详细介绍；第 4 节开展系统性方法实验评测，集中横向对比讨论各类方法的优缺点；第 5 节介绍解决图像缺损修复问题目前所仍面临的主要挑战和核心技术难点；第 6 节总结全文.

2 相关基础知识概述

本节将对相关基础知识进行简明介绍，包括深度神经网络、基于图像自身信息的缺损修复技术、评价指标、评测数据集等内容.

2.1 深度神经网络

深度神经网络是近年来出现并获得迅速发展的一种颠覆性机器学习知识模型[40]. 在结构组成上，相对于传统神经网络，也就是俗称的浅网络，深度神经网络有着明显增加的纵深，以及呈几何级数级扩容的神经元数目. 尽管在理论依据上还相对缺乏，但不可否认在实际表现方面，深度神经网络几乎在所有领域完全碾压其他机器学习经典模型，如 SVM[41]和 Boosting[42]等. 尤其在一般性图像建模这些涉及大规模复杂数据的问题上，深度神经网络的表现几无对手[43].

在具体结构上，深度神经网络通常由输入层、输出层以及部署在二者之间数目不等的隐含层组成，每层中有数目不等的多个神经元. 通过多个隐含层的叠加，网络可以较好实现对大规模复杂信息的高层级抽象建模. 通过设计搭配不同类型的输出层结构，深度网络能够基于隐含层中建模的复杂知识信息进行分类、识别、定位、分割等各类型决策操作. 在本文主要涉及的计算机视觉领域，目前被研究和应用最多的是卷积深度神经网络[44]. 卷积是一种具有一般意义的运算形式，通过采用搭配不同参数的卷积模版，绝大多数传统图像处理操作均可通过卷积实现，如平滑、锐化、采样等.

近年来，深度神经网络在高层级计算机视觉问题上取得一系列突出成绩，成为目前最突出的高级视觉知识建模方法. 卷积网络 AlexNet[45]、VGGNet[46]、GoogLeNet[47]、ResNet[48]等的提出，使得人们在 ImageNet[49]等超大规模实际图像数据库上进行高精度图像分类成为可能. 以 GAN[50]为代表的生成对抗类网络模型，如 CGAN[51]、DCGAN[52]、Stack-GAN[53]等，以博弈方式有效提升合成内容的灵活性，成为目前图像合成领域最令人瞩目的研究方向之一.

2.2 基于图像自身信息的缺损修复技术

如第 1 节所述，解决图像缺损修复问题的核心关键点在补偿知识的获取. 应对该问题，除了采用本文将在后续章节进行重点介绍的深度神经网络之外，从图像中未损坏的其它区域获取补偿信息，基于已知内容填补未知，也是一种可行的思路. 实际上，在深度神经网络这种突出的复杂信息建模技术出现之前，基本上这也是唯一可行的一般性图像缺损修复思路.

基于图像本身信息进行缺损修复，重点在如何实现补偿信息的有效选择和融合. 目前，该类技术主要可以被划分为两类：基于补丁（Patch-based）方法[54-60]和基于像素扩散（Diffusion-based）方法[61-71]. 基于补丁方法的主要思路是在图像中未缺损区域中搜索能够匹配缺损区域的补丁，而后将补丁填充至缺损区域；基于像素扩散方法通过对无缺损区域与缺损区域交界处的像素进行扩散，将无缺损区域的有效图像信息平滑地传递至缺损区域.

基于自身信息实现缺损修复的方法思路清晰、含义明确，同时也易于理解和应用，曾是解决该类问题最有效的方式. 然而，由于缺乏对图像高层次语义的理解，也并未构建广泛收集图像一般性信息的补偿知识模型，该类技术在细节精细程度、补偿灵活性、与实际图像的吻合度等方面均存在较为明

显的缺陷. 表现在往往只能局部复制移动图像内容, 而无法实现真正意义上的缺损补偿和灵活修复, 对于结构相对复杂, 本身冗余度低的图像, 通常难以达到理想修复效果. 基于此, 本文对该类技术不展开, 只重点总结介绍深度神经网络类图像缺损修复技术.

2.3 评价指标

由于属于相对较新的研究课题, 对于缺损图像修复, 尚无行业公认的专有量化评价指标. 对于相关方法的评价, 目前主要是基于直观定性观察, 或者图像整体质量评价指标进行. 对于后者, 近年来最常用的是峰值信噪比 PSNR（Peak Signal to Noise Ratio）[72] 和结构相似度 SSIM（Structural Similarity）[73]. 这两种指标都是通过对比处理后所获得图像 I_x 与目标图像 I_y 之间的差别来对方法效果进行评价. 其具体定义方式如下①:

$$PSNR(I_x, I_y) = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE(I_x, I_y)}\right), \quad (1)$$

$$SSIM(I_x, I_y) = \frac{(2\mu_{I_x}\mu_{I_y} + c_1)(2\sigma_{I_xI_y} + c_2)}{(\mu_{I_x}^2 + \mu_{I_y}^2 + c_1)(\sigma_{I_x}^2 + \sigma_{I_y}^2 + c_2)}, \quad (2)$$

其中

$$MSE(I_x, I_y) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}(I_x(i,j) - I_y(i,j))^2, \quad (3)$$

为图像 I_x 与图像 I_y 的均方误差; MAX_i 表示输入图像格式允许的灰度值上限, 对灰度图像, 其值为 255; μ_{I_x} 和 μ_{I_y} 分别表示图像 I_x 与图像 I_y 的像素均值; σ_{I_x} 与 σ_{I_y} 分别为图像 I_x 与图像 I_y 的方差; $\sigma_{I_xI_y}$ 表示协方差; c_1, c_2 为常数; H 与 W 分别为图像的高度和宽度.

2.4 评测数据集

标准的图像数据集是科学评价图像缺损修复算法的基础, 我们对图像修复领域常用的数据集进行了分类和汇总. 经过调研发现, 图像修复领域常用的数据集主要分为场景数据集, 物体数据集与一般性数据集三大类（详细数据集信息参见表 1）.

表 1 图像修复领域常用标准评测数据集

数据集名称	类型	分辨率	规模	网址	数据集被使用情况
Places2[74]	场景	256×256	10M	http://places2.csail.mit.edu/	[8],[14],[15],[16],[17],[18],[19],[26],[30],[32],[33],[34],[35],[36],[37]
Paris StreetView[75]		936×537	15K	https://github.com/pathak22/context-encoder	[5],[7],[8],[14],[21],[25],[30],[37]
Cityscapes[76]		2048×1024	25K	https://www.cityscapes-dataset.com/	[21],[22],[29]
Facades[77]		250×250~1024×1024	0.6K	http://cmp.felk.cvut.cz/~tylecr1/facade/	[19],[21]
SVHN[78]		32×32	600K	http://ufldl.stanford.edu/housenumbers/	[27]
CelebA[79]	物体图像	178×218	202K	http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html	[8],[16],[17],[23],[24],[26],[27],[30],[32],[33],[37]
CelebA-HQ[80]		1024×1024	30K	https://github.com/willylulu/celeba-hq-modified	[8],[15],[18],[19],[25],[31],[32],[33],[35]
PIPA[81]		50×50~1024×1024	51K	https://www.sarahelizabethmalinak.com/photo-albums/	[9]
MegaFace[82]		32×32~2048×2048	4.7M	http://megaface.cs.washington.edu/	[23]
Stanford Cars[83]		100×100~4000×4000	16K	https://ai.stanford.edu/~jkrause/cars/car_dataset.html	[27]
Helen Face[84]		480×480~3500×3500	2K	http://www.ifp.illinois.edu/~vuongle2/helen/	[24],[29]
ImageNet[49]	一般性图像	full resolution	14M	http://www.image-net.org/	[5],[6],[7],[8],[10],[15],[18],[25],[32],[35],[36]
DTD[85]		300×300~640×640	5K	https://www.robots.ox.ac.uk/~vgg/data/dtd/	[19],[32]

① 对于如 RGB 等多通道图像, 以各通道上指标的均值为图像整体度量指标。

场景数据集通常由在一种或多种特定场景中拍摄的真实图像组成. 比较有代表性的有包含 400 余不同场景图像的 Places2[74]多场景数据集；Paris StreetView[75]巴黎街景数据集，该数据集主要从 Google StreetView 数据集[86]抽取图像构成；大规模城市街景数据集 Cityscapes[76]；建筑物正面图像数据集 Facades[77]；街景门牌数据集 SVHN[78]. 场景数据集的特点是图像数据为客观世界中的真实场景. 由于真实场景的类型众多，且大多背景复杂，包含多种物体，该类数据集对于图像修复研究有一定难度.

物体数据集一般指数据集中的图像样本属于某一特定类型物体，如人脸等. 常见的此类数据集包括：大规模人脸数据集 CelebA[79]；CelebA 的高质量人脸图像版本 CelebA-HQ[80]；大型人脸图像数据集 Megaface[82]；包含各种外观变化下人脸图像的 Helen Face[84]数据集；包含人类在各种环境、活动、事件中的姿势与体态的 PIPA[81]人体姿态数据集，其图像主要从公共相册数据集 Flickr[87]中收集获得；汽车图像数据集 Stanford Cars[83]. 物体数据集与场景数据集不同，其背景的复杂度往往较低，且图像中的前景物体一般占据较大范围比例. 另外，物体数据集图像样本中的前景一般只包含一个物体. 因此可以看出，物体数据集的整体修复难度一般低于场景数据集.

一般性数据集指由大量不同类别图像构成的综合性数据集，比较典型的有多物体多场景综合数据集 ImageNet[49]，覆盖 47 种不同类别物体的纹理图像数据集 DTD[85]. 一般性数据集由于不受场景与物体类型的制约，内容更为多样，组成成分也更为复杂. 因此，在此类数据集上开展图像修复研究的整体难度也更高.

3　研究现状与分析

本节将对目前在图像缺损修复方面取得最突出表现的深度神经网络类方法进行系统性论述. 如图 2 所示，我们将主流方法依据核心网络架构的不同具体细分为五大类，即 Context-Encoder 类、U-Net 类、CGAN 类、DCGAN 类和 StackGAN 类.

3.1　Context-Encoder类方法

Context-Encoder[5]是一种无监督视觉特征学习网络架构. 在图像缺损修复问题中，上下文信息对于修复结果至关重要，Context-Encoder 网络中的上下文编码器能够有效利用待修复区域周围的局部信息与整张图像的全局信息，生成与原图像相匹配的信息. 以 Context-Encoder 为基础的图像缺损修复方法在近年来成为一种主流方式.

典型的 Context-Encoder 网络结构如图 3 所示，主要由编码器和解码器两部分组成. 编码器与解码器之间的全连接通道有效降低了网络参数的数量. Context-Encoder 不仅能够利用缺损区域周围的图像信息，还能利用整幅图像中视觉结构的语义信息. 在训练阶段，Context-Encoder 引入重建损失（Reconstruction Loss），如公式（4）所示，以获取缺失区域整体结构与上下文的一致性. 同时，Context-Encoder 也引入了对抗损失（Adversarial Loss），如公式（5）所示，来提升生成图像的灵活性. 方法的整体损失是上述重建损失与对抗损失的加权组合，如公式（6）所示.

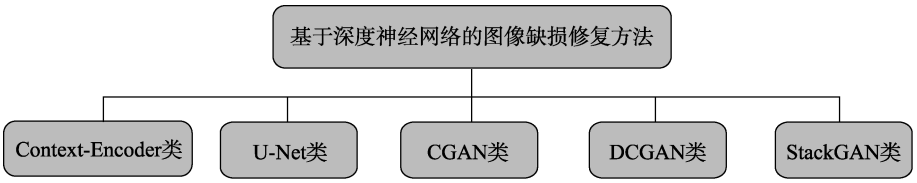

图 2　基于深度神经网络的图像缺损修复方法划分

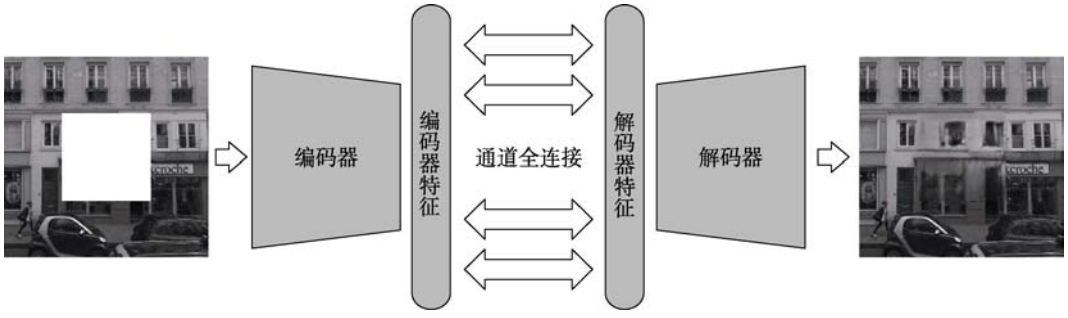

图 3　Context-Encoder 网络架构

$$\mathcal{L}_{rec}(I)=\| \widehat{M}\odot(I-F((1-\widehat{M})\odot I))\|_2^2, \quad (4)$$

$$\mathcal{L}_{adv}=\max_D \mathbb{E}_{I\in\chi}[\log(D(I))+\log(1-D(F((1-\widehat{M})\odot I)))], \quad (5)$$

$$L=\lambda_{rec}\mathcal{L}_{rec}+\lambda_{adv}\mathcal{L}_{adv}, \quad (6)$$

其中，I 代表输入图像，$\widehat{M}$ 为图像缺失区域的 mask，$\odot$ 代表矩阵元素对位乘，F 表示网络输出，χ代表真实样本分布，D 为鉴别器，λ_{rec} 与λ_{adv} 分别为重建损失与对抗损失的加权系数.

Liao 等[6]在 Context-Encoder 的基础上提出了基于边缘的上下文解码器. 该模型首先提取图像的边缘信息，然后利用全卷积网络（Fully Convolutional Network）来恢复缺失区域的边缘信息，最后将恢复的边缘信息与不完整图像用 Context-Encoder 进行修复. Yang 等人[7]将 Context-Encoder 中所有的 ReLU 层和 Leaky ReLU 层替换为 ELU 层，使网络在训练时能够处理较大的负面响应，保证了网络训练的稳定性. Wang 等人[8]提出的网络架构中包含三个平行的分支网络，每一个分支网络在下采样时使用不同大小的感受野，能够提取不同层级的图像特征，克服了单个网络中只能提取同一层级特征的问题. Sun 等[9]设计提出了一种专门修复人脸图像的方法，具体包括两个步骤：第一步先根据图像的上下文信息在合适的位置生成人脸特征点，第二部根据第一步的结果补全缺失的人脸图像. 两部分的生成器都应用 Context-Encoder 架构. Vo 等人[10]在 Context-Encoder 框架基础上提出了结构损失（Structural Loss），这是一种像素重建损失与特征重建损失的线性组合，以进一步提升方法精度.

3.2 U-Net类方法

U-Net[88]网络由 FCN[89]网络进化而来，最初应用于图像分割领域. U-Net 网络之所以被应用于图像缺损修复问题，是由于其具有独特的特征融合方式：U-Net 可以将不同尺度的特征在通道维度拼接在一起，其最后一个上采样输出的特征是由来自第一个卷积模块输出的特征与上一个上采样输出特征融合得来. 这种独特的特征融合方式将图像中的低层次特征与高层次特征深度融合. 在图像缺损修复领域，特征融合方式对图像缺损区域的修复至关重要，不仅需要利用颜色、边缘等低层次特征，也同时要结合结构、语义等高层次特征. 基于此，U-Net 结构在图像缺损修复领域有着广泛应用[11-19]. 如图 4 所示，U-Net 网络结构主要包括两部分：第一部分的主要作用为特征提取，网络架构与 VGG 网络架构[13]类似，网络层中每经过一个池化层就是一个尺度；第二部分为上采样部分，每上采样一次，其输出与特征提取部分对应融合. U-Net 网络没有全连接层，且卷积的过程中采用 valid 方式，保证结果均是从没有缺失的上下文特征中获得.

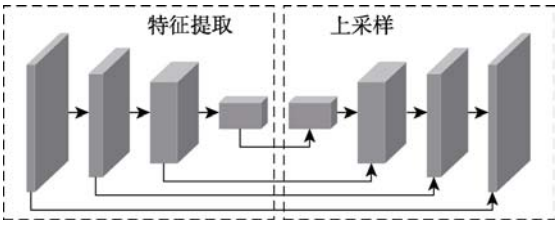

图 4 U-Net 网络架构

Yan 等人[14]在 U-Net 架构上增加特殊的移位连接层，提出了 Shift-Net 结构，能够填充具有复杂结构和精细纹理的任意形状缺损区域. 移位连接层能够链接已知区域的编码器特征和待修复区域的解码器特征，通过合理的语义和精细的细节纹理产生高质量的修复结果. Liu 等人[15]将 U-Net 中所有卷积层替换为部分卷积（Partial Convolution），并在解码阶段使用最近邻上采样，移位连接层分别将编码器和解码器的特征和掩码连接在一起，作为下一个部分卷积层的输入. 最后一个部分卷积层的输入为原始缺失图像和原始掩码. Wang 等人[16]在 U-Net 中加入多尺度注意力模块，不仅计算图像低层细节的相似程度，也同时计算图像高层语义相似性. Hong 等人[17]在 U-Net 解码器的最后几层中嵌入了多个融合模块，设计提出 DFNet 方法. 每个融合模块从相应特征图产生不同分辨率的修复结果. 在损失方面，对于高层特征，添加感知损失 $\mathcal{L}_p^k$（Perceptual Loss）与风格损失 $\mathcal{L}_s^k$（Style Loss）. 感知损失为原始图像与修复图像的相似特征之间的 L_1 距离，如公式（7）所示；风格损失为原始图像与修复图像对应的格拉姆矩阵的 L_1 距离，如公式（8）所示.

$$\mathcal{L}_p^k=\sum_{j\in J}\| \phi_j(I_k)-\phi_j(\hat{I}_k)\|_1, \quad (7)$$

$$\mathcal{L}_s^k=\sum_{j\in J}\| G_j^\phi(I_k)-G_j^\phi(\hat{I}_k)\|_1, \quad (8)$$

其中，I_k 为待修复图像，$\hat{I}_k$ 为被修复图像，$\phi_j(\cdot)$ 为网络层第 j 层的输出，是一个大小为 $C_j\times H_j\times W_j$ 的特征图，其中 C_j，H_j，W_j 分别代表特征图的通道数，高度与宽度；$G_j^\phi(\cdot)$ 表示格拉姆矩阵，其中每一个元素为

$$G_j^\phi(I)_{c,c'}=\frac{1}{C_jH_jW_j}\sum_{h=1}^{H_j}\sum_{w=1}^{W_j}\phi_j(I)_{h,w,c}\phi_j(I)_{h,w,c'}, \quad (9)$$

其中，c 和 c' 代表元素在矩阵中的位置坐标. 感知

损失与风格损失的加入，使缺损图像在修复过程中受到更多约束. 通过嵌入融合模块与这些约束，使网络能将修复内容与现有内容实现平滑融合，达到较好修复效果.

Xiao 等人[18]受 NIN（Network in Network）和 Inception Module 的启发，将 Inception Module 嵌入 U-Net. Inception Module 的主要思想是，卷积网络中经过优化的局部稀疏结构可以被一系列容易获得的密集子结构近似和替代，从而在保持网络稀疏性的同时提高计算性能. Zeng 等人[19]在 U-Net 结构之上，提出了金字塔上下文编码网络 PEN-Net. 它可以对来自全分辨率输入的上下文语义进行编码，并将学习到的语义特征解码为图像信息用于修复缺损内容. 在训练过程中，编码器从高级语义特征图中通过注意力逐步学习区域之间的关联性，然后通过注意力将区域之间的关联性转移到低级特征图中，确保图像修复的视觉和语义的一致性. Zhang 等人[20]将多个 U-Net 网络作为生成器，逐步修复图像中的缺失内容.

3.3　CGAN类方法

条件生成对抗网络 CGAN[51]是经典生成对抗网络 GAN 的条件约束版本. GAN 是近年来在图像生成领域出现的标志性方法，通过博弈机制的引入，其生成模式不再直接受限于训练目标图像，在生成内容灵活性方面的表现显著优于其它方法. 而图像缺损修复问题，在本质上是缺损内容的生成问题，因此 GAN 类方法在该问题上有着普遍应用.

由于对模型的约束较少，经典 GAN 对生成内容的整体控制力和定制力偏弱，限制了方法的适用范围和应用价值. 针对该问题，Mirza 等人[51]提出了能够施加额外约束要求的条件生成对抗网络 CGAN，网络架构如图 5 所示，目标函数如公式（10）. 与 GAN 不同的地方在于，它的生成器与鉴别器的输入分别多了一个条件，使其能够按照特定条件生成所需图像.

$$\min_G \max_D V(D,G) = E_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}\left[\log\left(D(\boldsymbol{x}\mid \boldsymbol{y})\right)\right] + E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\left[\log\left(1-D\left(G(\boldsymbol{z}\mid \boldsymbol{y})\right)\right)\right]. \quad (10)$$

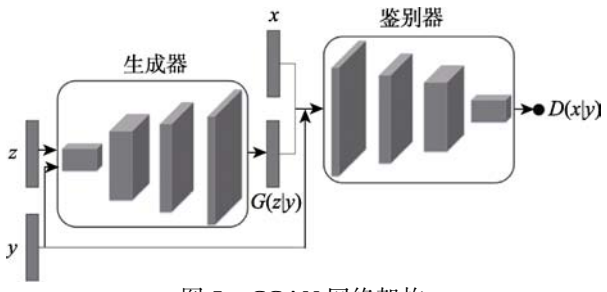

图 5　CGAN 网络架构

近年来以 CGAN 为核心的图像缺损修复方法取得一系列突出进展. Isola 等人[21]设计提出一种 CGAN 方法应用的一般性框架，即 PIX2PIX. 由于基础 CGAN 往往局限于生成低分辨率的图像，Wang 等人[22]提出了基于 CGAN 的具有多尺度生成器和鉴别器的网络架构 HRISSM，可以生成 2048×1024 的高分辨率图像. Dolhansky 等人[23]基于 CGAN 使用隐藏在参考图像中的示例信息，生成高质量眼睛修复图像. 其网络使用一组对应的训练集图像 X_i 和 r_i 来进行对抗训练，如图 6 所示，其中，X_i 为生成器 G 的输出，r_i 为参考图像，Z_i 为缺损图像，D 为鉴别器. Liao 等人[24]基于 CGAN 提出了协作生成对抗网络，该网络由三个分支网络组成，即人脸特征点生成网络，图像修复网络，图像分割网络. 每一个分支网络输出的结果都与输入一起作为鉴别器的输入. 该框架通过将额外信息嵌入到其它任务中，从而诱导性地改进主要生成任务. Zheng 等人[25]基于 CGAN 架构提出双分支网络 Pluralistic，其中一个分支为重建路径，根据图像中已知区域信息获得缺失区域信息的先验分布；另一分支为生成路径，结合上一分支中获得的条件先验信息，生成缺损部分内容.

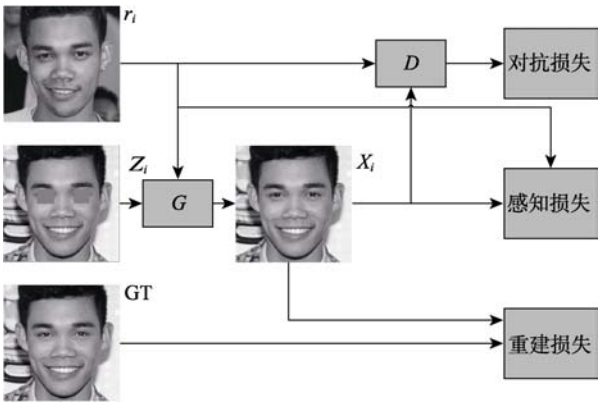

图 6　文献[23]提出的眼睛修复网络结构

3.4　DCGAN类方法

DCGAN[52]是 CNN[90]与 GAN[50]深层次融合的产物. 相对于经典 GAN，DCGAN 在特征提取方式方面的升级，有效提高了网络训练的稳定性与图像生成质量. 图像缺损修复问题是根据已有图像信息生成缺损部分的图像内容，实际上就是求从已有图像信息到完整图像的映射关系. 这种映射关系在图像中的表示就是特征之间的映射. 在图像中学习更丰富的特征表示是求解映射关系的关键，DCGAN 在该方面表现突出. 具体实现上，DCGAN 对卷积神经网络结构做了一些改变，主要包括：（1）去掉所有的池化层，生成器中使用微步幅度卷积代替池

化层，鉴别器中使用步幅卷积代替池化层；（2）生成器和鉴别器中均使用批量归一化；（3）去掉全连接层，使网络变为全卷积网络；（4）生成器中使用 ReLU 作为激活函数，最后一层使用 tanh；（5）鉴别器中使用 Leaky ReLU 作为激活函数. DCGAN 的网络结构图如图 7 所示，生成器的输入是一个 100 维的噪声，中间通过 4 层卷积层，每通过一个卷积层通道数减半，长宽扩大一倍，最终产生 64×64×3 大小的图像输出；鉴别器可以看成是生成器取反. 值得一提的是，DCGAN 中生成器中采用的卷积为微步幅度卷积，而不是反卷积，两者的差别如图 8 所示，反卷积在整个输入矩阵周围补 0，而微步幅度卷积会把输入矩阵拆开，在每一个像素点的周围补 0.

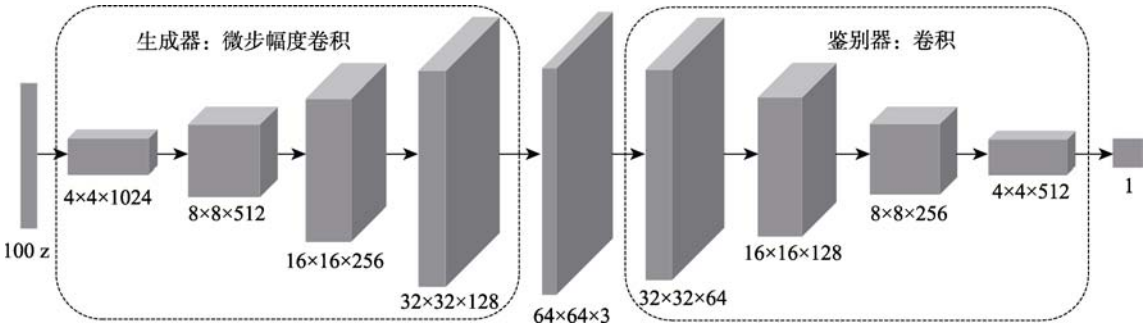

图 7 DCGAN 网络结构

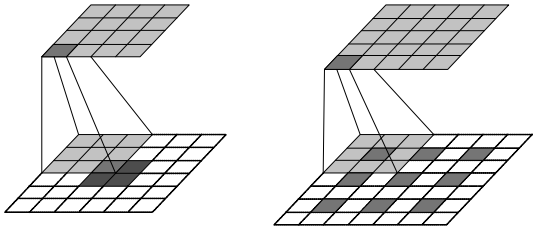

图 8 反卷积与 DCGAN 中使用的微步幅度卷积，其中左边为反卷积，右边为微步幅度卷积.

Iizuka 等人[26]将 DCGAN 应用于图像修复领域，设计提出 GL 方法，引入全局和局部上下文鉴别器进行修复图像鉴别，其中全局鉴别器判断整张图像的合理性，局部鉴别器判断被修复局部区域的合理性. Yeh 等人[27]提出基于 DCGAN 的网络模型，并使用完好无损的图像训练生成器，通过对现有数据的学习来生成图像中的缺失内容. 在测试阶段，生成模型通过搜索潜在图像中与受损图像的最近编码，并与未缺损部分的内容信息相匹配然后生成缺失的内容信息. Banerjee 等人[28]提出了一个基于 DCGAN 的多尺度的生成对抗网络，能够从一个单一的面部图像生成让人产生幻觉的虚拟现实环境和背景像素. 该模型由多个 DCGAN 模块级联而成，每一个模块生成不同分辨率的图像，下一个模块的输入图像是由上一个模块生成的图像经过上采样将分辨率扩大两倍得到的. Song 等人[29]首次将语义分割信息引入到图像修复中，将图像中的类间差异和类内差异分离开来，使得语义不同区域之间的边界恢复得更清晰，语义一致的区域内纹理更精确. 提出的模型分为两个模块，第一个模块为分割预测网络，基于全卷积神经网络，主要作用是对缺失区域的分割标签进行预测；第二个模块为分割引导模块，分割引导模块基于 DCGAN，根据分割结果生成图像内容信息. Nazeri 等人[30]提出了一种结合边缘先验信息的图像修复方法 EdgeConnect. 其中，生成模型先生成完整边缘信息，再结合缺损图像一同生成修复图像. 结构上，该模型为二阶段生成对抗网络，两个阶段均基于 DCGAN. 第一阶段的生成器输入为缺损图像，基于 HED 检测得到的边缘图像以及缺损区域掩码，输出为完整的边缘图像；第二阶段修复生成器的输入为缺损图像和完整边缘图. Jo 等人[31]提出了基于 DCGAN 的 SC-FEGAN，该网络具有高度的交互性，能够在图像大面积缺失的情况下进行图像修复. 网络的生成器架构为 U-Net 结构，输出层的激活函数采用 tanh，判别器的架构为 SN-PatchGAN.

3.5 StackGAN类方法

前文介绍的 CGAN 类方法在整体约束模型上有了长足进步，但由于网络结构相对简单，学习到的图像特征不如 U-Net 与 DCGAN 丰富，其所修复的图像在清晰程度方面有一定的不足. 近年来在图像合成领域取得突出进展，通过层级式结构的引入有效提升所生成图像清晰程度的 StackGAN 类方法，有效弥补了该方面的不足.

原始 StackGAN[53]为一个两阶段 Coarse-to-Fine 生成对抗网络，网络结构如图 9 所示. 第一阶段以

给定的文本描述为基础生成对象的原始形状和基本颜色，并根据随机噪声矢量生成背景；第二阶段根据第一部分的结果作为输入校正第一阶段生成的低分辨率图像中的缺陷，并根据文本描述来细化生成对象中的纹理细节，从而生成高分辨率的逼真图像.

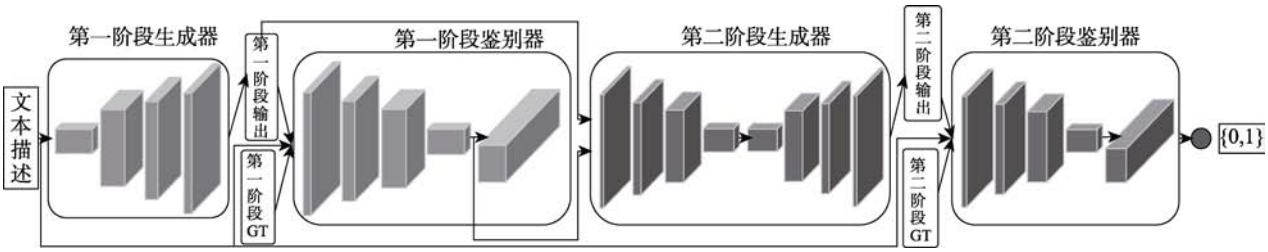

图 9　StackGAN 网络结构

在 StackGAN 架构体系下，Yu 等人[32]提出一个由粗到精的两阶段网络架构 GIICA，第一阶段用重建损失进行训练，第二阶段用重建损失以及对抗损失进行训练. 第二阶段的编码器可以比第一阶段学习到更好的特征表示，能够获得更完整的场景. Yu 等人[33]在文献[32]的基础上，将卷积换为门控卷积，设计提出 GConv 方法. 门控卷积能够为图像中的每个通道和空间位置进行动态特征选择，可以根据背景、掩码、人机交互勾勒的草图学习选择特征图，以生成更好的修复结果. Xiong 等人[34]提出了由三个模块级联而成的图像修复网络. 其第一个模块为不完整轮廓检测模块，该模块应用 Deepcut 网络，产生前景不完整物体的边缘；第二个模块和第三个模块的网络架构基于 StackGAN，先生成粗略的边缘图和粗略的内容信息，而后再进行精细化生成. Sagong 等[35]提出基于 StackGAN 的 PEPSI 网络，该网络将两级级联网络统一为一个单级编解码器网络. 该网络由一个单一的共享编码器和两个并行解码器组成. 两个并行解码器一个为粗修复路径，一个为精修复路径. 粗修复路径从编码器产生的特征映射产生粗略修复结果，精修复路径首先通过 CAM 重建特征图，然后解码器通过重建的特征图生成更高质量的修复结果. Shin 等人[36]在文献[35]的基础上提出了轻量级的 PEPSI 模型 Diet-PEPSI，利用少量参数聚合全局背景信息，缩短了训练时间. Liu 等人[37]提出基于 StackGAN 的两阶段网络结构，并在第二阶段中增加相关注意力网络层，以提升图像修复的清晰程度.

4　方法性能分析

为全面评估方法性能，对比方法表现，本节将在大规模公开数据集上对所述方法展开系统性综合实验.

4.1　实验配置

本文全部实验基于 Ubuntu18.04 操作系统进行，统一采用 Intel E5-2620 CPU，32GB 内存，Tesla K80 GPU（12GB 显存）硬件平台.

考虑到如图 1 所示的中心区域缺损和随机区域缺损是目前相关研究重点关注的问题，我们的实验将在这两类缺损图像上全面展开. 由于缺损区域面积显著影响方法修复效果，为全面度量方法能力，我们将在不同缺损面积下，分别开展系统性实验. 具体配置为：在中心区域缺损上，实验 10%、15%、20%、25%四种缺损比例；在随机区域缺损上，实验 10%、20%、30%、40%四种缺损比例. 注：两类比例配置不同的原因在于，中心缺损定位在图像的中心位置，因此影响比随机缺损更大，所以我们对缺损比例做了适当缩减.

众所周知，深度神经网络类方法涉及的配置参数众多，在没有源代码的情况下，很难准确复现原有算法. 因此，为客观评价方法水平，我们将仅对提供公开代码的方法进行实验. 在实验过程中，最大限度保持原方法配置不改变，确保客观呈现原有方法水平.

具体实验时，每组实验均由定量度量、定性展示，以及资源消耗三部分内容组成，全面展示评估方法表现. 其中，定量实验基于 PSNR 和 SSIM 两项指标在全部测试集图像上的平均表现进行考察；定性方面，为全面展示方法性能，我们没有刻意选择效果最好或者最差的图像，而是随机选取部分结果进行展示，重在体现方法的一般性效果水平；资源消耗则按照在所有相关实验上的平均表现进行记录统计.

4.2　数据集

本文实验基于目前图像缺损修复领域最常用的两个公开数据集 CelebA 和 Places2 进行. 如 2.4 节所述，CelebA 是由人脸图像组成的著名单一物

体图像数据集,而 Places2 则是包含数百个实际场景的综合性场景数据集.

在 CelebA 上，随机选择 1,000 幅图像组成测试集,训练集由剩余 201,599 幅图像组成;Places2 数据集上，从其验证子集中随机选择 1,000 幅图像作为测试集，另使用该数据集测试子集全体 328,500 幅图像作为训练集. 本文所有实验均在相同训练集和测试集配置下进行.

4.3 实验分析

4.3.1 Context-Encoder 类方法

表 2 展示了 Context-Encoder 类方法在中心区域缺损图像上的定量修复表现①. 从表中可以看到，CE[5]方法和 GMCNN[8]方法各在 8 项实验中占优，反映出两类方法的精度表现无确定性优劣关系.

图 10 展示了方法的相应定性表现. 从图中可以看出，两种方法的表现与定量结果基本一致，未有一种方法能够达到普遍意义上的更优表现. CE 方法在 CelebA 上的表现稍好，GMCNN 则在 Places2 数据集上有一定优势.

表 3 展示了两种方法的资源消耗情况. 可以看出，在训练时间、测试时间、使用内存和显存方面,CE 方法全面占优. 因此可以得出结论,CE 是一种在运算速度和资源消耗方面更为优秀的方法. 本质上,这是由于 CE 的结构相对于 GMCNN 明显更为精简：在特征编码阶段 CE 方法只有 1 个模块，而 GMCNN 则有 3 个.

表 2 Context-Encoder 类方法在中心区域缺损图像上的修复表现

数据集	CelebA								Places2							
评价指标	*PSNR*				*SSIM*				*PSNR*				*SSIM*			
遮挡比例	10%	15%	20%	25%	10%	15%	20%	25%	10%	15%	20%	25%	10%	15%	20%	25%
CE[5]	26.20	**29.10**	**28.15**	**30.75**	0.910	**0.928**	**0.911**	**0.931**	20.20	21.49	**21.22**	**21.16**	0.780	0.777	0.759	0.751
GMCNN[8]	**28.26**	25.86	24.31	22.63	**0.942**	0.915	0.891	0.856	**24.13**	**22.12**	20.35	18.79	**0.915**	**0.881**	**0.840**	**0.793**

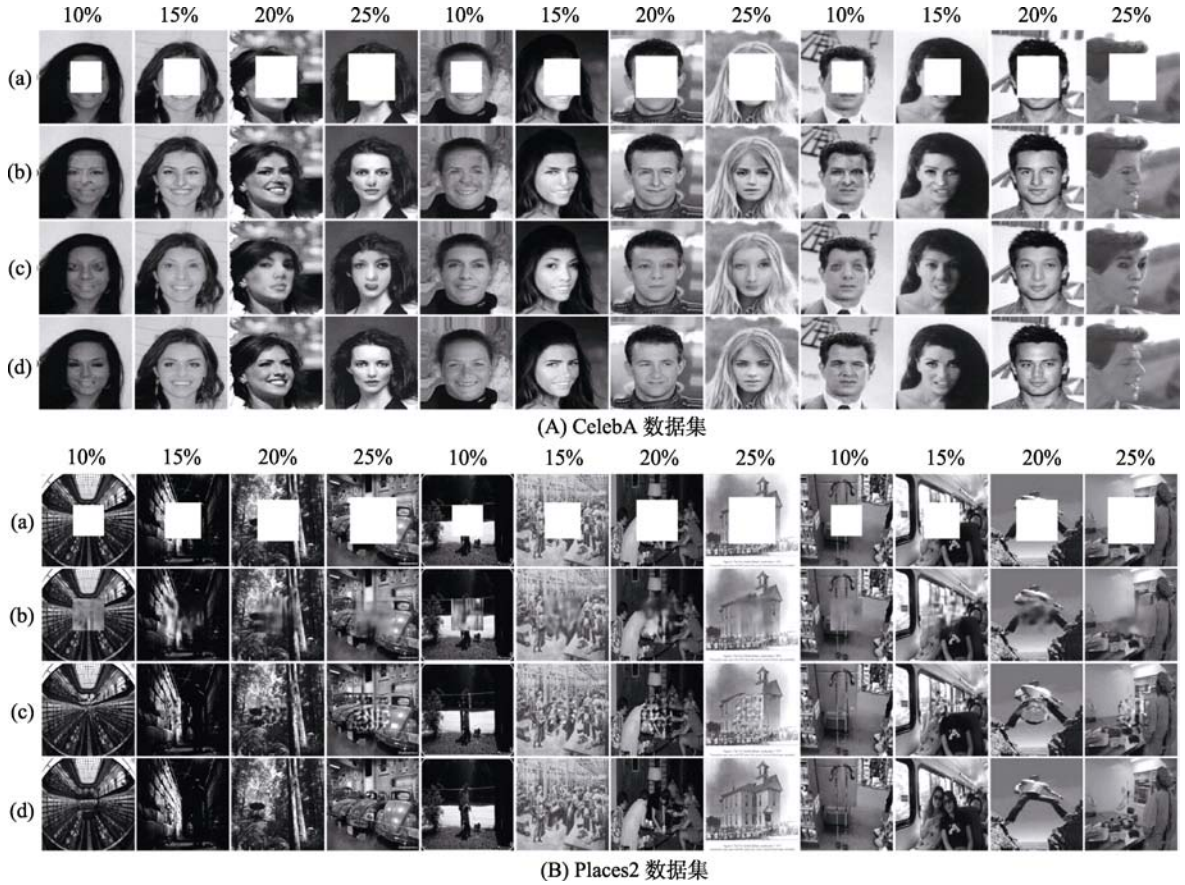

图 10 Context-Encoder 类方法在中心区域缺损图像上的修复效果：(a) 输入图像，(b) CE 方法修复结果，(c) GMCNN 方法修复结果，(d) Ground Truth.

① 本文实验使用的 CE Pytorch 版本源代码中没有随机区域缺损训练源码，因此该轮实验中没有随机区域缺损部分。

表 3　Context-Encoder 类方法资源消耗情况

资源	训练时间(h)	单幅测试时间(s)	使用内存(GB)	使用显存(GB)	开源代码网址
数据集	CelebA/Places2				
CE[5]	**106.6/133.3**	**0.089/0.095**	**1.52/1.56**	**1.88/2.18**	https://github.com/BoyuanJiang/context_encoder_pytorch
GMCNN[8]	170.1/220.0	0.222/0.236	**1.52**/1.87	4.96/6.36	https://github.com/shepnerd/inpainting_gmcnn

4.3.2　U-Net 类方法

表 4 与表 5 分别展示了 U-Net 类方法在中心区域缺损和随机区域缺损图像上的定量实验表现. 从表中可以看出，DFNet 方法[17]在 PSNR 与 SSIM 指标上的表现全面占优（全部 32 项测试中，DFNet 有 29 项取得最优成绩）. 可见，DFNet 引入的感知损失和风格损失，以及融合模块嵌入策略，对于 U-Net 类方法的性能提升有显著作用.

图 11 与图 12 展示了相应定性实验表现. 从图中可以看出，与定量表现一致，DFNet 方法能够取得明显更好的修复效果.

表 6 展示了 U-Net 类方法的资源消耗情况. 可以看到，DFNet 在最重要的测试时间方面，有明显优势，能够用更少的运算时间完成图像修复. 但在显存使用方面，相对于 Shift-Net[14]方法明显偏高.

图 11　U-Net 类方法在中心区域缺损图像上的修复效果：(a) 输入图像，(b) Shift-Net 方法修复结果，(c) DFNet 方法修复结果，(d) PEN-Net 方法修复结果，(e) Ground Truth.

图 12 U-Net 类方法在随机区域缺损图像上的修复效果：(a) 输入图像，(b) Shift-Net 方法修复结果，(c) DFNet 方法修复结果，(d) PEN-Net 方法修复结果，(e) Ground Truth.

表 4 U-Net 类方法在中心区域缺损图像上的修复表现

数据集	CelebA								Places2							
评价指标	*PSNR*				*SSIM*				*PSNR*				*SSIM*			
遮挡比例	10%	15%	20%	25%	10%	15%	20%	25%	10%	15%	20%	25%	10%	15%	20%	25%
Shift-Net[14]	24.23	22.53	22.51	22.82	0.903	0.856	0.876	**0.865**	24.97	23.44	22.19	20.98	0.902	0.873	0.839	0.802
DFNet[17]	**30.90**	**29.21**	**27.25**	**25.17**	**0.934**	**0.920**	**0.900**	0.860	**25.44**	**23.76**	**22.35**	**21.10**	**0.916**	**0.885**	**0.849**	**0.808**
PEN-Net[19]	26.29	24.87	23.57	21.70	0.892	0.873	0.852	0.817	22.79	21.99	21.12	20.22	0.817	0.794	0.765	0.729

表 5 U-Net 类方法在随机区域缺损图像上的修复表现

数据集	CelebA								Places2							
评价指标	*PSNR*				*SSIM*				*PSNR*				*SSIM*			
遮挡比例	10%	20%	30%	40%	10%	20%	30%	40%	10%	20%	30%	40%	10%	20%	30%	40%
Shift-Net[14]	**32.20**	29.10	26.54	23.62	**0.943**	0.913	0.870	0.808	28.91	25.11	22.69	20.72	0.930	0.868	0.792	0.712
DFNet[17]	31.56	**30.94**	**29.17**	**26.96**	0.942	**0.920**	**0.900**	**0.864**	**29.82**	**26.07**	**23.64**	**21.63**	**0.945**	**0.889**	**0.825**	**0.754**
PEN-Net[19]	27.03	23.92	21.59	20.04	0.882	0.819	0.759	0.713	23.39	20.93	19.04	17.78	0.792	0.703	0.622	0.564

表 6　U-Net 类方法资源消耗情况

资源	训练时间(h)	单幅测试时间(s)	使用内存(GB)	使用显存(GB)	开源代码网址
数据集	CelebA/Places2				
Shift-Net[14]	**49.9**/317.0	0.210/0.360	**1.67**/2.27	**1.43**/**1.67**	https://github.com/Zhaoyi-Yan/Shift-Net_pytorch
DFNet[17]	153.8/**208.3**	**0.060**/**0.065**	1.75/**1.96**	4.38/4.58	https://github.com/hughplay/DFNet
PEN-Net[19]	157.0/220.5	0.154/0.160	2.63/5.84	5.51/7.70	https://github.com/researchmm/PEN-Net-for-Inpainting

4.3.3　CGAN 类方法

表 7 与表 8 分别展示了 CGAN 类方法在中心区域缺损和随机区域缺损图像上的定量图像修复表现. 通过结果可以看出，Pluralistic 方法[25]有明显精度优势：在 90.6%的实验中能够取得更好表现. 同时可以看到，Pluralistic 方法的量化表现与缺损区域面积成反比，与一般性预期相符，说明方法的稳定性较好. 上述实验表现说明，Pluralistic 方法提出的双分支网络架构，是一种有效的 CGAN 类方法实现策略.

图 13、图 14 展示了相关定性结果. 可以看到，两种方法的性能与定量表现基本一致. 但是 Pluralistic 方法在修复人脸图像时，在多幅图像上获得的结果与 Ground Truth 人脸有较大差异，反映出此类方法的精度水平仍需进一步提升.

表 9 展示了资源消耗情况. 在训练时间、使用内存和显存方面，Pluralistic 方法有一定优势，但在测试时间方面弱于 PIX2PIX[21]. 这是由于 PIX2PIX 方法采用了结构最为简单的单生成器结合单判别器体系架构.

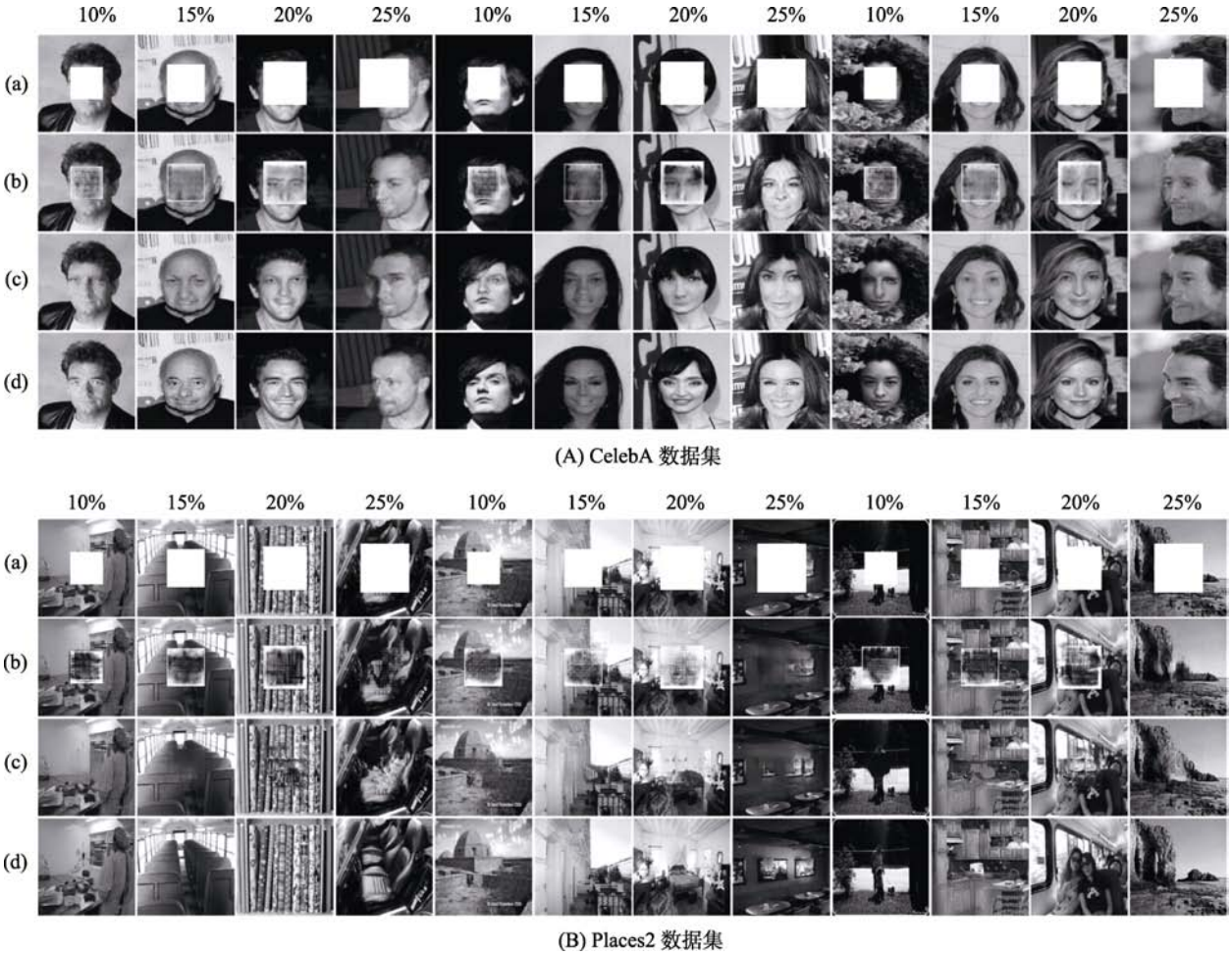

图 13　CGAN 类方法在中心区域缺损图像上的修复效果：(a) 输入图像，(b) PIX2PIX 方法修复结果，(c) Pluralistic 方法修复结果，(d) Ground Truth.

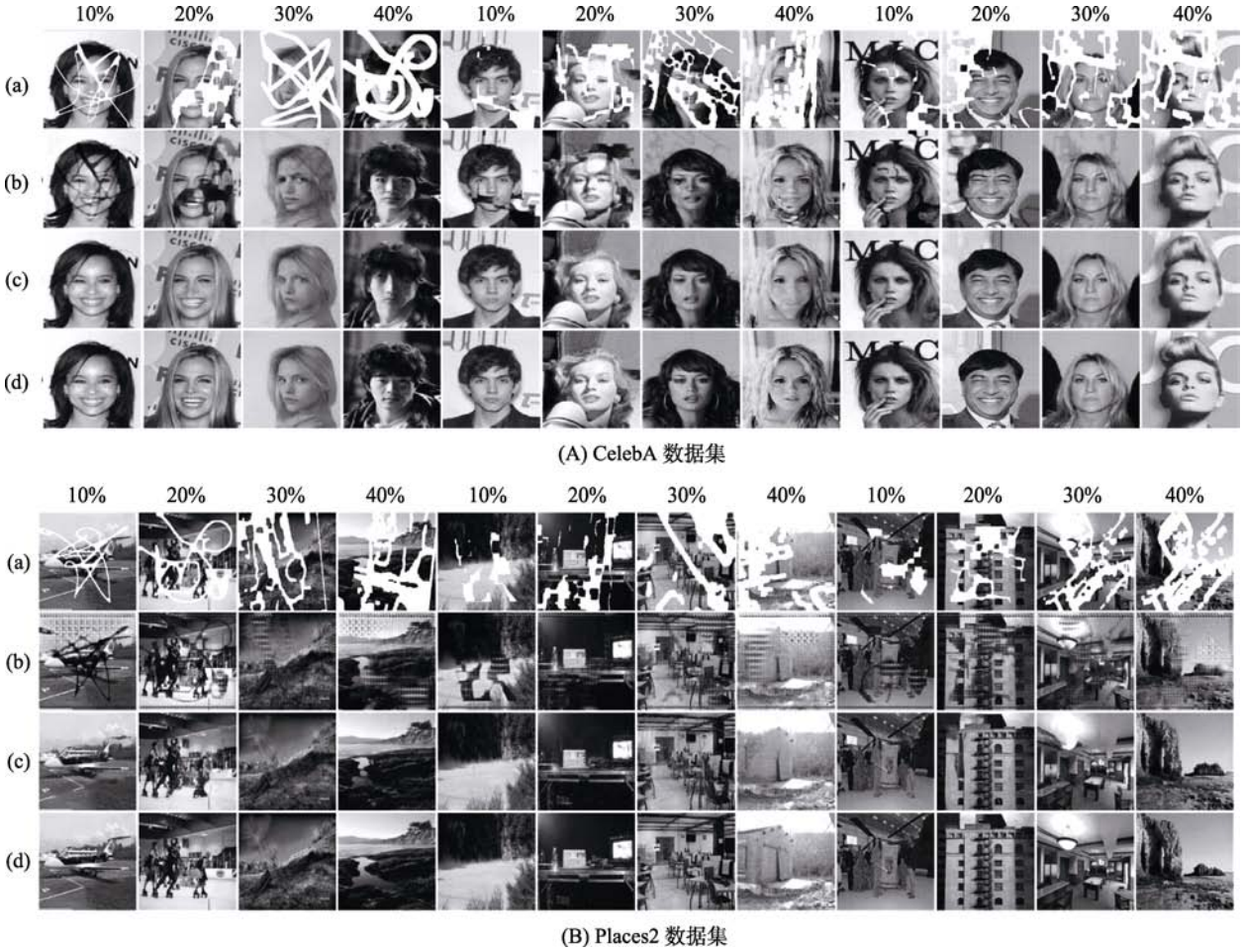

图 14 CGAN 类方法在随机区域缺损图像上的修复效果：(a) 输入图像，(b) PIX2PIX 方法修复结果，(c) Pluralistic 方法修复结果，(d) Ground Truth.

表 7 CGAN 类方法在中心区域缺损图像上的修复表现

数据集	CelebA								Places2							
评价指标	*PSNR*				*SSIM*				*PSNR*				*SSIM*			
遮挡比例	10%	15%	20%	25%	10%	15%	20%	25%	10%	15%	20%	25%	10%	15%	20%	25%
PIX2PIX[21]	20.59	21.17	17.05	**22.79**	0.856	0.844	0.804	0.834	16.97	17.08	15.05	**19.60**	0.840	0.814	0.771	0.784
Pluralistic[25]	**28.14**	**25.13**	**23.50**	21.47	**0.941**	**0.912**	**0.883**	**0.842**	**23.81**	**22.20**	**20.72**	19.32	**0.907**	**0.874**	**0.835**	**0.795**

表 8 CGAN 类方法在随机区域缺损图像上的修复表现

数据集	CelebA								Places2							
评价指标	*PSNR*				*SSIM*				*PSNR*				*SSIM*			
遮挡比例	10%	20%	30%	40%	10%	20%	30%	40%	10%	20%	30%	40%	10%	20%	30%	40%
PIX2PIX[21]	17.67	18.40	23.37	**23.68**	0.792	0.767	0.811	0.786	15.97	16.02	17.11	16.92	0.679	0.639	0.594	0.532
Plurallistic[25]	**32.95**	**28.88**	**26.19**	23.42	**0.957**	**0.919**	**0.874**	**0.814**	**28.11**	**24.73**	**22.46**	**20.38**	**0.926**	**0.864**	**0.798**	**0.723**

表 9 CGAN 类方法资源消耗情况

资源	训练时间(h)	单幅测试时间(s)	使用内存(GB)	使用显存(GB)	开源代码网址
数据集	CelebA/Places2				
PIX2PIX[21]	229.2/466.7	**0.021**/**0.021**	2.94/3.06	11.40/11.40	https://github.com/affinelayer/pix2pix-tensorflow
Pluralistic[25]	**108.0**/**336.0**	0.535/0.670	**2.16**/**2.31**	**6.20**/**6.25**	https://github.com/lyndonzheng/Pluralistic-Inpainting

4.3.4　DCGAN 类方法

表 10、表 11 展示了 DCGAN 类方法在中心区域缺损和随机区域缺损图像上的定量实验结果. 可以看到，该系列实验体现出鲜明的数据集决定性：在 CelebA 数据集上，EdgeConnect[30]方法全面占优，而在 Places2 集合上，GL[26]方法则全面领先. 出现该现象的原因，主要在于 EdgeConnect 方法注重边缘信息，而人脸图像上的边缘信息较为单一和明确，所以能够起到较好的指示作用. 但在场景数据集 Places2 上，边缘信息较为复杂，核心边缘信息不突出，对图像修复的指示作用有所降低.

图 15、图 16 展示了定性实验结果. 图中反映的精度信息与定量实验基本一致. 但需要注意的是，图 16 中部分图像容易引发人类视觉系统的错觉，误以为 EdgeConnect 的表现优于 GL. 这是因为偏重边缘信息的 EdgeConnect 往往会在细节纹理上接近 Ground Truth 图像（因为纹理的特征与细小的边缘信息直接相关），但在整体结构上却无法体现原物体特点，造成细节相近但结构缺失. 如(B)图最后一列所示，虽然 EdgeConnect 的修复结果在纹理上与 Ground Truth 图相近，但仔细对比可以发现，原图中重要的白色书籍、水龙头等内容在修复图像中消失了.

从表 12 展示的资源消耗情况看，EdgeConnect 方法在测试时间和训练时间上的表现优于 GL 方法，但需要消耗更多的显存资源.

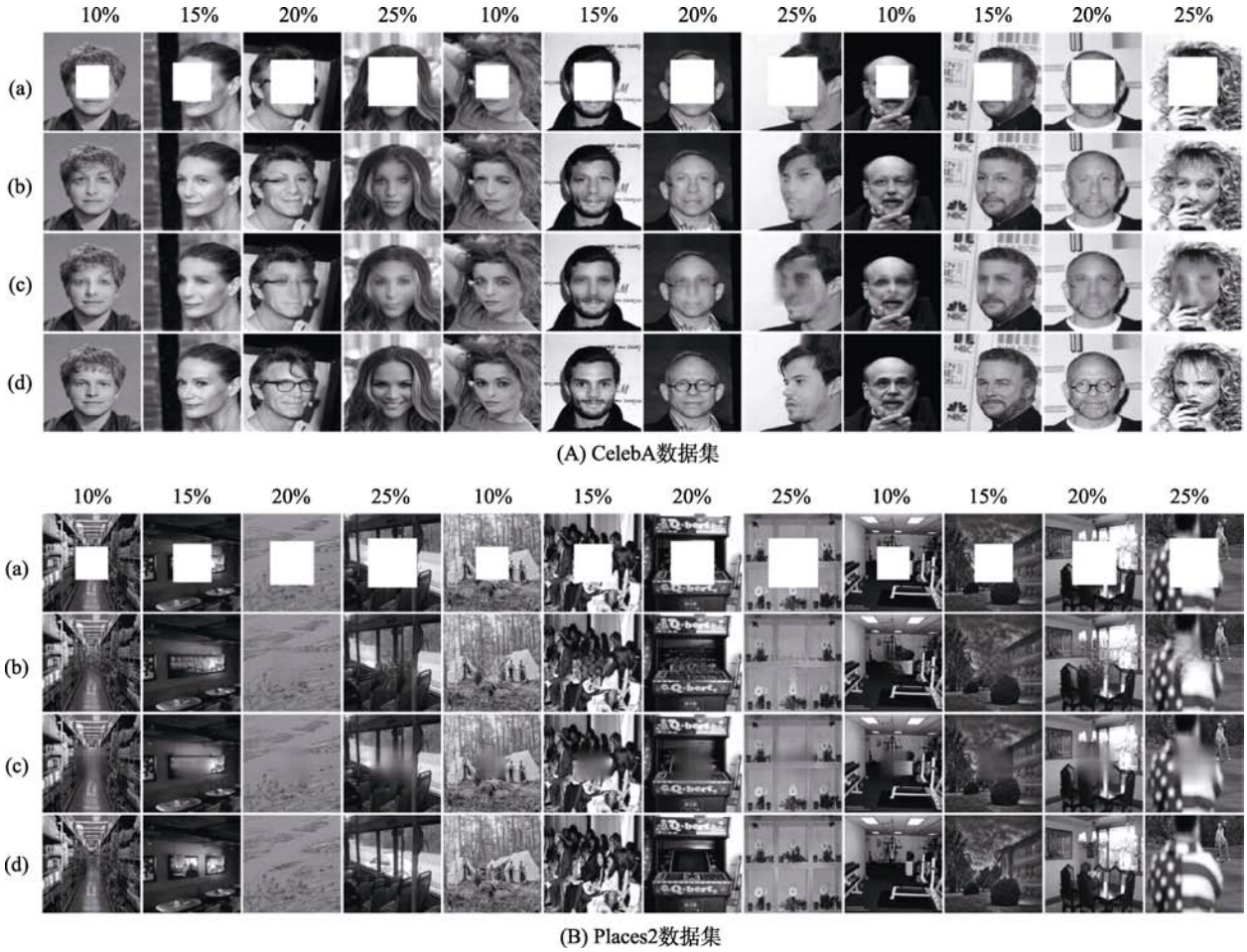

图 15　DCGAN 类方法在中心区域缺损图像上的修复效果：(a) 输入图像，(b) EdgeConnect 方法修复结果，(c) GL 方法修复结果，(d) Ground Truth.

表 10　DCGAN 类方法在中心区域缺损图像上的修复表现

数据集	CelebA								Places2							
评价指标	*PSNR*				*SSIM*				*PSNR*				*SSIM*			
遮挡比例	10%	15%	20%	25%	10%	15%	20%	25%	10%	15%	20%	25%	10%	15%	20%	25%
EdgeConnect[29]	**29.44**	**25.98**	**24.73**	**22.52**	**0.951**	**0.919**	**0.894**	**0.853**	25.24	23.58	22.17	20.94	0.916	0.887	0.851	0.809
GL[25]	26.39	24.49	23.60	21.77	0.847	0.826	0.809	0.781	**27.02**	**25.19**	**23.65**	**22.28**	**0.932**	**0.906**	**0.875**	**0.840**

表 11 DCGAN 类方法在随机区域缺损图像上的修复表现

数据集	CelebA								Places2							
评价指标	*PSNR*				*SSIM*				*PSNR*				*SSIM*			
遮挡比例	10%	20%	30%	40%	10%	20%	30%	40%	10%	20%	30%	40%	10%	20%	30%	40%
EdgeConnect[29]	**35.29**	**30.87**	**28.14**	**25.43**	**0.974**	**0.944**	**0.908**	**0.860**	30.04	26.02	23.74	21.66	0.950	0.893	0.833	0.764
GL[25]	27.65	26.20	24.25	21.44	0.863	0.840	0.804	0.744	**34.76**	**28.32**	**25.10**	**22.18**	**0.976**	**0.936**	**0.884**	**0.812**

表 12 DCGAN 类方法资源消耗情况

资源	训练时间(h)	单幅测试时间(s)	使用内存(GB)	使用显存(GB)	开源代码网址
数据集	CelebA/Places2				
EdgeConnect[29]	**165.8**/**204.0**	**0.48**/**0.54**	**2.10**/2.11	9.53/9.56	https://github.com/knazeri/edge-connect
GL[25]	236.0/218.0	4.76/1.05	2.79/**1.64**	**4.22**/**4.51**	https://github.com/shinseung428/GlobalLocalImageCompletion_TF

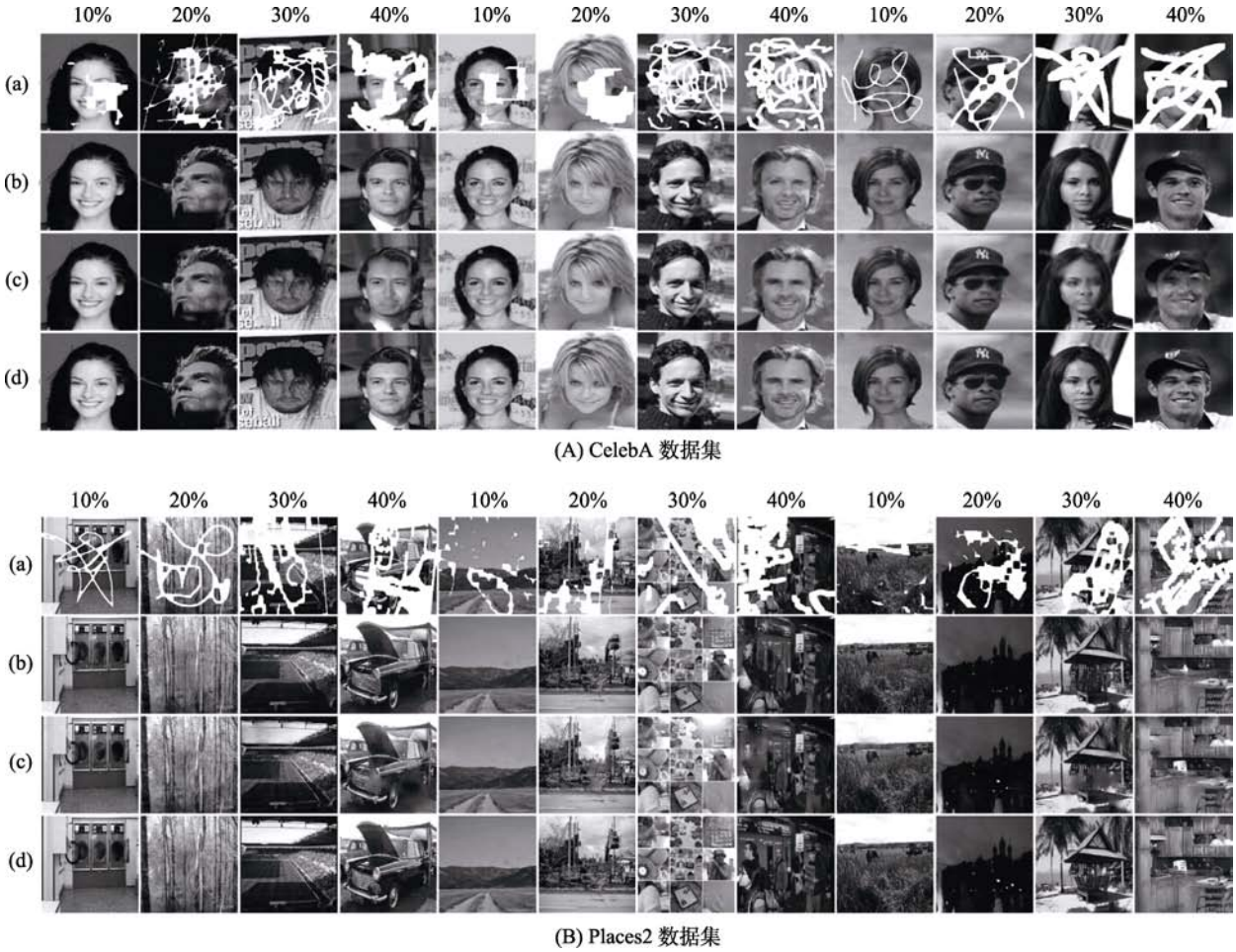

图 16 DCGAN 类方法在随机区域缺损图像上的修复效果：(a) 输入图像，(b) EdgeConnect 方法修复结果，(c) GL 方法修复结果，(d) Ground Truth.

4.3.5 StackGAN 类方法

方法的定量修复表现集中展示在表 13 和表 14 中. 可以看到，在约 75%的实验中，GConv[33]方法的表现更好. 图 17、图 18 展示了相应定性实验结果，GConv 方法同样展示出良好性能，在多幅测试图像上取得良好效果，验证了 StackGAN 类方法提出的层级式结构能够有效帮助图像修复工作.

从表 15 展示的资源消耗情况看，GConv 方法的测试运行速度更快，且消耗的显存资源更少.

4.3.6 综合表现

在所有方法的综合表现方面，呈现明显多元化现象：

在 CelebA 数据集上，针对中心区域缺损问题，表现最好的是 Context-Encoder 类的 CE 方法，在 5/8

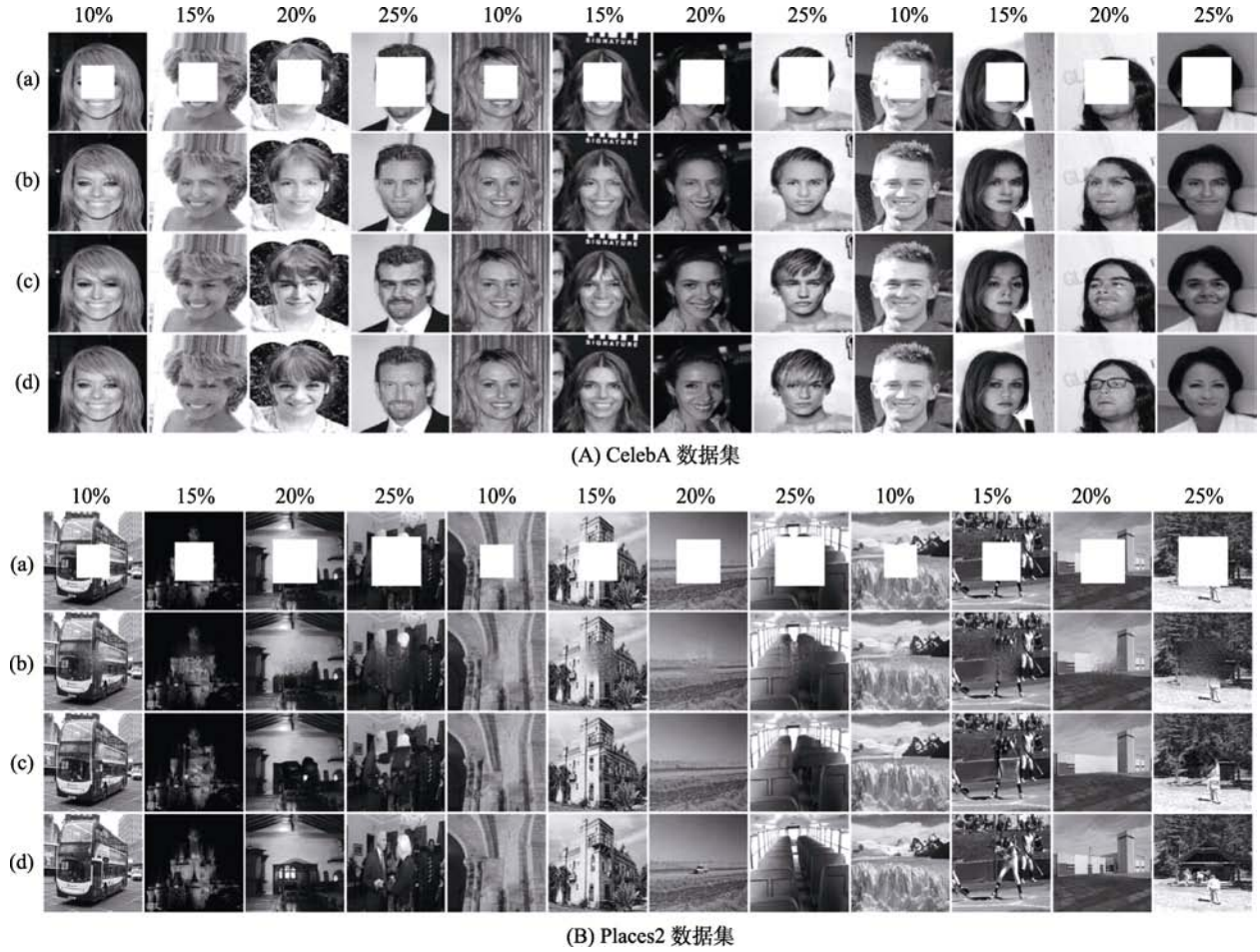

图 17　StackGAN 类方法在中心区域缺损图像上的修复效果：(a) 输入图像，(b) CSA 方法修复结果，(c) GConv 方法修复结果，(d) Ground Truth.

表 13　StackGAN 类方法在中心区域缺损图像上的修复表现

数据集	CelebA								Places2							
评价指标	*PSNR*				*SSIM*				*PSNR*				*SSIM*			
遮挡比例	10%	15%	20%	25%	10%	15%	20%	25%	10%	15%	20%	25%	10%	15%	20%	25%
CSA[36]	28.53	26.29	24.67	23.22	0.907	0.886	0.865	0.834	**24.55**	**23.20**	**21.99**	**20.81**	0.908	0.879	0.844	0.802
GConv[32]	**29.75**	**27.33**	**25.04**	**23.23**	**0.954**	**0.928**	**0.900**	**0.862**	23.94	22.18	20.77	19.53	**0.913**	**0.882**	**0.846**	**0.805**

表 14　StackGAN 类方法在随机区域缺损图像上的修复表现

数据集	CelebA								Places2							
评价指标	*PSNR*				*SSIM*				*PSNR*				*SSIM*			
遮挡比例	10%	20%	30%	40%	10%	20%	30%	40%	10%	20%	30%	40%	10%	20%	30%	40%
CSA[36]	30.36	28.23	26.45	24.47	0.914	0.891	0.862	0.819	29.04	**25.76**	**23.63**	**21.70**	0.936	0.886	0.831	**0.765**
GConv[32]	**34.86**	**30.06**	**27.34**	**24.89**	**0.972**	**0.937**	**0.897**	**0.849**	**30.27**	25.63	23.00	20.86	**0.953**	**0.897**	**0.834**	**0.765**

表 15　StackGAN 类方法资源消耗情况

资源	训练时间(h)	单幅测试时间(s)	使用内存(GB)	使用显存(GB)	开源代码网址
数据集	CelebA/Places2				
CSA[36]	**42.2**/**342.0**	2.580/3.360	**2.05**/**2.03**	8.29/8.46	https://github.com/KumapowerLIU/CSA-inpainting
GConv[32]	1132/566	**0.234**/**0.235**	2.98/2.93	**6.15**/**6.15**	https://github.com/JiahuiYu/generative_inpainting

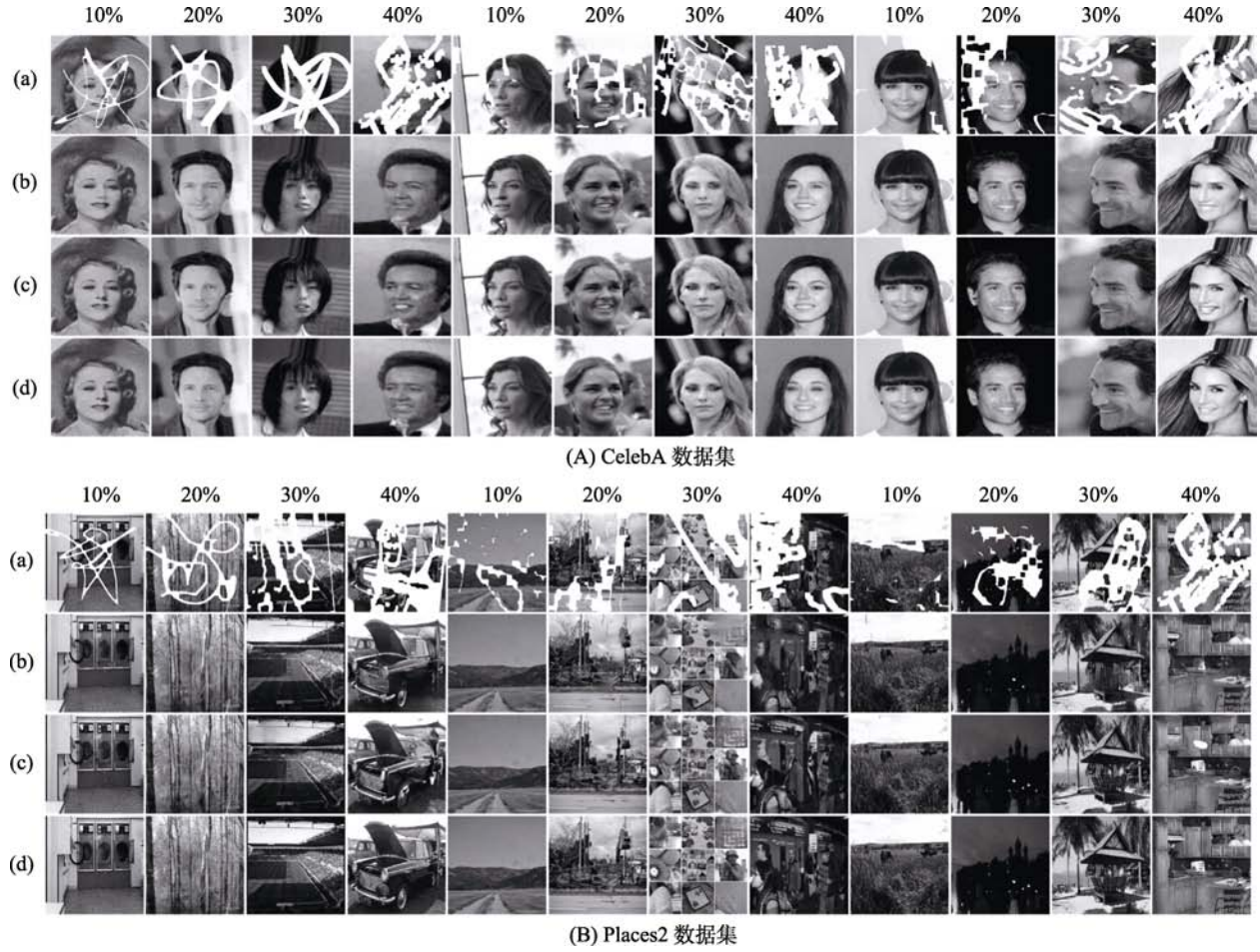

图 18 StackGAN 类方法在随机区域缺损图像上的修复效果：(a) 输入图像，(b) CSA 方法修复结果，(c) GConv 方法修复结果，(d) Ground Truth.

项量化指标上取得最优表现；其次是 U-Net 类的 DFNet 方法，在 2/8 项指标上取得最优表现. 针对随机缺损问题，表现最好的是 U-Net 类的 DFNet 和 DCGAN 类的 EdgeConnect，分别在 4/8 项指标上现最佳. 在运算速度方面，CGAN 类的 PIX2PIX 方法最快，单幅图像平均处理时间 0.021 秒. 而训练用时最少的，则是 StackGAN 类的 CSA，时长 42.2 小时. 显存占用方面，表现最好的是 U-Net 类的 Shift-Net，平均占用 1.43GB. 内存方面，占用最少的是 Context-Encoder 类的 CE 和 GMCNN，均为 1.52GB.

在 Places2 数据集上，面对中心区域缺损和随机缺损图像，表现最好的都是 DCGAN 类的 GL 方法在，其在全部 16 项量化指标上均取得最优表现. 运算速度方面，CGAN 类的 PIX2PIX 方法达到单幅图像平均 0.021 秒的处理速度，表现最优. 训练时间上，Context-Encoder 类的 CE 方法在 133.3 小时内完成训练过程，耗时最短. 显存和内存消耗上，与 CelebA 数据集上的情况类似，U-Net 类的 Shift-Net 和 Context-Encoder 类的 CE 方法消耗最低，分别为 1.67GB 和 1.56GB.

从上述实验表现看，数据集的差异、图像缺损类型的不同等因素均会对最优方法的选择产生直接影响. 因此，目前在图像缺损修复领域并没有一种能够稳定实现最优表现的突出方法. 而且从定性效果看，即使是在量化表现方面最优秀的方法，在我们随机展示的实验图像上也并非全部能够达到十分理想的效果，反映出整个领域还存在比较大的提升空间.

5 所面临的问题和挑战

毋庸置疑，缺损图像修复问题有着明确的应用背景和突出的实际价值，是计算机视觉、图像处理、机器学习等领域不可或缺的重要研究课题. 而深度神经网络技术的出现和发展，有效促进了该领域整体技术水平的全面提升. 但深度网络出现的时间并不长，大面积应用于图像修复问题也仅短暂的数年时间，整体技术水平并不成熟，仍存在一定提升空间. 除了前文已经提到的精度问题，目前所面临的

其它问题和挑战可能包括：

（1）缺损修复既涉及高层语义知识，又离不开低层像素信息，只有将两部分信息高结构化融合，才能逼近人类视觉系统的图像修复水平. 目前的深度神经网络架构，在该方面仍缺乏妥善解决方案.

（2）受限于庞大的网络结构所带来的巨额运算开销，目前深度神经网络类方法往往只能针对分辨率较低图像进行处理. 本文研究的方法中，只有 HRISSM[22]、GIICA[32]和 GConv[33]能够对 512×512 以上图像进行修复. 高分辨率图像的缺损修复问题目前亟待解决.

（3）在图像修复领域，目前大量方法均需要提前获知缺损区域，因此并不是完全意义上的端到端解决方案. 设计能够彻底摆脱人工干预和前期预分析处理的全自动图像修复深度神经网络架构仍是当前重要研究课题.

（4）模型的泛化能力仍存在较大提升空间. 在目前的相关研究中，方法还是仍然必须在相关度较高的训练集上进行训练才能在对应测试集上取得较好效果. 仍然缺少一种一般意义上的“最小泛化训练数据集”，即只要在该数据集上充分训练，就可以在大多数图像修复问题上取得理想效果.

（5）专门性度量指标仍然缺乏. 从本文实验部分可以看出，目前采用比较普遍的 PSNR 和 SSIM 两项指标虽然对修复效果有一定指示作用，但与人类的直观视觉体验尚存差距. 设计专门的图像修复质量度量指标和方法，势在必行.

6　总　结

本文对基于深度神经网络的图像缺损修复技术展开系统研究. 将目前主流技术划分为 Context-Encoder 类、U-Net 类、CGAN 类、DCGAN 类以及 StackGAN 类共五个类别，对每个类别的核心思路和技术特点进行了归纳和分析介绍. 对于公开代码的 11 种方法进行了系统性评测实验和性能对比. 最后，对于目前相关研究中存在的问题和挑战进行了介绍和阐述.

参　考　文　献

[1] Huang Y, Sun S, Duan X, et al. A study on deep neural networks framework//Proceedings of the IEEE Conference on Advanced Information Management, Communicates, Electronic and Automation Control Conference. Xi'an, China, 2016: 1519-1522

[2] Wang R, Li G and Chu D. Capsules encoder and capsgan for image inpainting//Proceedings of the International Conference on Artificial Intelligence and Advanced Manufacturing. Dublin Ireland, 2019: 325-328

[3] Xiang P, Wang L, Cheng J, et al. A deep network architecture for image inpainting//Proceedings of the IEEE International Conference on Computer and Communications. Chengdu, China, 2017: 1851-1856

[4] Su Y, Liu T, Liu K, et al. Image inpainting for random areas using dense context features//Proceedings of the IEEE International Conference on Image Processing. Taipei, China, 2019: 4679-4683

[5] Pathak D, Krahenbuhl P, Donahue J, et al. Context encoders: Feature learning by inpainting//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 2536-2544

[6] Liao L, Hu R, Xiao J, et al. Edge-aware context encoder for image inpainting//Proceedings of the International Conference on Acoustics, Speech and Signal Processing. Calgary, Canada, 2018: 3156-3160

[7] Yang C, Lu X, Lin Z, et al. High-resolution image inpainting using multi-scale neural patch synthesis//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 6721-6729

[8] Wang Y, Tao X, Qi X, et al. Image inpainting via generative multi-column convolutional neural networks//Proceedings of the Conference and Workshop on Neural Information Processing Systems. Montreal, Canada, 2018: 331-340

[9] Sun Q, Ma L, Joon Oh S, et al. Natural and effective obfuscation by head inpainting//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 5050-5059

[10] Vo H V, Duong N Q K, Perez P. Structural inpainting. arXiv preprint arXiv: 1803.10348, 2018

[11] Liu Y, Qi N, Zhu Q, et al. CR-U-Net: Cascaded U-Net with residual mapping for liver segmentation in CT images// Proceedings of the IEEE International Conference on Visual Communications and Image Processing. Sydney, Australia, 2019: 1-4

[12] Han Y, Ye J C, Framing U-Net via deep convolutional framelets: application to sparse-view CT. IEEE Transactions on Medical Imaging, 2018, 37(6): 1418-1429

[13] Zhang Z, Liu Q and Wang Y. Road extraction by deep residual U-Net. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5): 749-753

[14] Yan Z, Li X, Li M, et al. Shift-Net: Image inpainting via deep feature rearrangement//Proceedings of the European Conference on Computer Vision. Munich, Germany, 2018: 1-17

[15] Liu G, Reda F A, Shih K J, et al. Image inpainting for irregular holes using partial convolutions//Proceedings of the European Conference on Computer Vision. Munich, Germany, 2018: 85-100

[16] Wang N, Li J, Zhang L, et al. MUSICAL: multi-scale image contextual attention learning for inpainting//Proceedings of the International Joint Conference on Artificial Intelligence. Macao, China, 2019

[17] Hong X, Xiong P, Ji R, et al. Deep fusion network for image completion. arXiv preprint arXiv: 1904.08060, 2019

[18] Xiao Q, Li G, Chen Q. Deep inception generative network for cognitive image inpainting. arXiv preprint arXiv: 1812.01458,

2018

[19] Zeng Y, Fu J, Chao H, et al. Learning pyramid-context encoder network for high-quality image inpainting//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. California, USA, 2019: 1486-1494

[20] Zhang H, Hu Z, Luo C, et al. Semantic image inpainting with progressive generative networks//Proceedings of the ACM International Conference on Multimedia. Seoul, Korea, 2018: 1939-1947

[21] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 1125-1134

[22] Wang T C, Liu M Y, Zhu J Y, et al. High-resolution image synthesis and semantic manipulation with conditional GANs//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 8798-8807

[23] Dolhansky B, Canton Ferrer C. Eye in-painting with exemplar generative adversarial networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 7902-7911

[24] Liao H, Funka-Lea G, Zheng Y, et al. Face completion with semantic knowledge and collaborative adversarial learning// Proceedings of the Asian Conference on Computer Vision. Perth, Australia, 2018: 382-397

[25] Zheng C, Cham T, Cai J, Pluralistic image completion// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 1438-1447

[26] Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion. ACM Transactions on Graphics, 2017, 36(4): 107

[27] Yeh R A, Chen C, Yian Lim T, et al. Semantic image inpainting with deep generative models//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 5485-5493

[28] Banerjee S, Scheirer W J, Bowyer K W, et al. On hallucinating context and background pixels from a face mask using multi-scale GANs. arXiv preprint arXiv: 1811.07104, 2018

[29] Song Y, Yang C, Shen Y, et al. SPG-Net: segmentation prediction and guidance network for image inpainting. arXiv preprint arXiv: 1805.03356, 2018

[30] Nazeri K, Ng E, Joseph T, et al. Edgeconnect: generative image inpainting with adversarial edge learning. arXiv preprint arXiv: 1901.00212, 2019

[31] Jo Y, Park J. SC-FEGAN: face editing generative adversarial network with user's sketch and color. arXiv preprint arXiv: 1902.06838, 2019

[32] Yu J, Lin Z, Yang J, et al. Generative image inpainting with contextual attention//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 5505-5514

[33] Yu J, Lin Z, Yang J, et al. Free-form image inpainting with gated convolution//Proceedings of the IEEE International Conference on Computer Vision. California, USA, 2019: 4471-4480

[34] Xiong W, Yu J, Lin Z, et al. Foreground-aware image inpainting//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. California, USA, 2019: 5840-5848

[35] Sagong M, Shin Y, Kim S, et al. PEPSI: fast image inpainting with parallel decoding network//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. California, USA, 2019: 11360-11368

[36] Shin Y G, Sagong M C, Yeo Y J, et al. PEPSI++: fast and lightweight network for image inpainting. arXiv preprint arXiv: 1905.09010, 2019

[37] Liu H, Jiang B, Xiao Y, et al. Coherent semantic attention for image inpainting//Proceedings of the IEEE International Conference on Computer Vision. Seoul, Korea, 2019: 4169-4178

[38] Elharrouss O, Almaadeed N, Al-Maadeed S, et al. Image inpainting: a review. Neural Process Letters, 2020, 51(2): 2007–2028

[39] Qiang Z, He L, Chen X, et al. Survey on deep learning image inpainting methods. Journal of Image and Graphics, 2019, 24(03): 447-463(in Chinese)
(强振平, 何丽波, 陈旭等. 深度学习图像修复方法综述. 中国图象图形学报, 2019, 24(03): 447-463.)

[40] Brahma P, Wu D and She Y, Why deep learning works: a manifold disentanglement perspective. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(10): 1997-2008

[41] Suykens J A. Support vector machines: a nonlinear modelling and control perspective. European Journal of Control, 2001, 7 (2-3):311-327

[42] Friedman J H. Greedy function approximation: a gradient boosting machine. Annals of Statistics, 2001, 29(5): 1189-1232

[43] Mishkin D, Sergievskiy N, Matas J. Systematic evaluation of convolution neural network advances on the imagenet. Computer Vision and Image Understanding, 2017, 161: 11-19

[44] Al-Saffar A A M, Tao H and Talab M A. Review of deep convolution neural network in image classification//Proceedings of the International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications. Jakarta, Indonesia, 2017: 26-31

[45] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks//Proceedings of the Conference and Workshop on Neural Information Processing Systems. Lake Tahoe, USA, 2012: 1097-1105

[46] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 1409.1556, 2014

[47] Szegedy C, Liu W, Jia Y et al. Going deeper with convolutions//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 1-9

[48] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 770-778

[49] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA, 2009: 248-255

[50] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets//Proceedings of the Conference and Workshop on Neural Information Processing Systems. Montreal, Canada, 2014: 2672-2680

[51] Mirza M, Osindero S. Conditional generative adversarial nets.

arXiv preprint arXiv: 1411.1784, 2014

[52] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv: 1511.06434, 2015

[53] Zhang H, Xu T, Li H, et al. StackGAN: text to photo-realistic image synthesis with stacked generative adversarial networks// Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 5907-5915

[54] Ružić T, Pižurica A. Context-aware patch-based image inpainting using Markov random field modeling. IEEE Transactions on Image Processing, 2014, 24(1): 444-456

[55] Demir U, Unal G. Patch-based image inpainting with generative adversarial networks. arXiv preprint arXiv: 1803.07422, 2018

[56] Newson A, Almansa A, Gousseau Y, et al. Non-local patch-based image inpainting. Image Processing on Line, 2017, 7: 373-385

[57] Guo Q, Gao S, Zhang X, et al. Patch-based image inpainting via two-stage low rank approximation. IEEE Transactions on Visualization and Computer Graphics, 2017, 24(6): 2023-2036

[58] Le Meur O, Gautier J, Guillemot C. Examplar-based inpainting based on local geometry//Proceedings of the IEEE International Conference on Image Processing. Brussels, Belgium, 2011: 3401-3404

[59] Ružić T, Pižurica A. Texture and color descriptors as a tool for context-aware patch-based image inpainting. The International Society for Optical Engineering, 2012, 8295(1): 52

[60] Bi X, Liu H, Lu G, et al. Exemplar-based inpainting under Boundary contraction constraints//Proceedings of the IEEE International Conference on Automation, Electronics and Electrical Engineering. Shenyang, China, 2018: 295-300

[61] He K, Sun J. Statistics of patch offsets for image completion// Proceedings of the European Conference on Computer Vision. Firenze, Italy, 2012: 16-29

[62] Sun J, Yuan L, Jia J, et al. Image completion with structure propagation. ACM Transactions on Graphics, 2005, 24(3): 861- 868

[63] Ružić T, Pižurica A. Context-aware patch-based image inpainting using markov random field modeling. IEEE Transactions on Image Processing, 2015, 24(1): 444-456

[64] He K, Sun J. Image completion approaches using the statistics of similar patches. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(12): 2423-2435

[65] Barnes C, Shechtman E, Finkelstein A, et al. PatchMatch: a randomized correspondence algorithm for structural image editing. ACM Transactions on Graphics, 2009, 28(3): 24

[66] Criminisi A, Pérez P, Toyama K. Region filling and object removal by exemplar-based image inpainting. IEEE Transactions on Image Processing, 2004, 13(9): 1200-1212

[67] Huang J B, Kang S B, Ahuja N, et al. Image completion using planar structure guidance. ACM Transactions on Graphics, 2014, 33(4): 129

[68] Muddala S M, Olsson R, Sjöström M. Spatio-temporal consistent depth-image-based rendering using layered depth image and inpainting. Eurasip Journal on Image and Video Processing, 2016, 2016(1): 1-19

[69] Isogawa M, Mikami D, Iwai D, et al. Mask optimization for image inpainting. IEEE Access, 2018(6): 69728-69741

[70] Liu J, Yang S, Fang Y, et al. Structure-guided image inpainting using homography transformation. IEEE Transactions on Multimedia, 2018, 20(12): 3252-3265

[71] Zeng J, Fu X, Leng L, et al. Image inpainting algorithm based on saliency map and gray entropy. Arabian Journal for Science and Engineering, 2019, 44(4): 3549-3558

[72] Johnson D H. Signal-to-noise ratio. Scholarpedia, 2006, 1(12): 2088

[73] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 2004, 13(4): 600-612

[74] Zhou B, Lapedriza A, Khosla A, et al. Places: a 10 million image database for scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(6): 1452-1464

[75] Doersch C, Singh S, Gupta A, et al. What makes Paris look like Paris? Communications of the ACM, 2015, 58(12): 103-110

[76] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 3213-3223

[77] Tylecek R, Sara R, Spatial pattern templates for recognition of objects with regular structure//Proceedings of the German Conference on Pattern Recognition. Saarbrücken, Germany, 2013: 364-374

[78] Yuval Netzer, Tao Wang, Adam Coates, et al. Reading digits in natural images with unsupervised feature learning//Proceedings of the Conference and Workshop on Neural Information Processing Systems. Granada, Spain, 2011

[79] Liu Z, Luo P, Wang X, et al. Deep learning face attributes in the wild//Proceedings of the IEEE International Conference on Computer vision. Santiago, Chile, 2015: 3730-3738

[80] Karras T, Aila T, Laine S, et al. Progressive growing of GANs for improved quality, stability, and variation. arXiv preprint arXiv: 1710.10196, 2017

[81] [81] Zhang N, Paluri M, Taigman Y, et al. Beyond frontal faces: improving person recognition using multiple cues//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 4804-4813

[82] Miller D, Brossard E, Seitz S, et al. Megaface: a million faces for recognition at scale. arXiv preprint arXiv: 1505.02108, 2015

[83] Krause J, Stark M, Deng J, et al. 3D object representations for fine-grained categorization//Proceedings of the IEEE International Conference on Computer Vision Workshops. Sydney, Australia, 2013: 554-561

[84] Le V, Brandt J, Lin Z, et al. Interactive facial feature localization//Proceedings of the European Conference on Computer Vision. Heidelberg, Germany, 2012: 679-692

[85] Cimpoi M, Maji S, Kokkinos I, et al. Describing Textures in the Wild//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 3606-3613

[86] Gronat P, Havlena M, Sivic J, et al. Building streetview datasets for place recognition and city reconstruction. Research Reports of CMP, Czech Technical University in Prague, 2011

[87] Huiskes M J, Lew M S. The MIR flickr retrieval evaluation// Proceedings of the ACM International Conference on Multimedia Information Retrieval. Vancouver, Canada, 2008: 39-43

[88] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation//Proceedings of the International Conference on Medical Image Computing and

Computer-assisted Intervention. Munich, Germany, 2015: 234-241

[89] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 3431-3440

[90] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278-2324

LI Yue-Long, Ph.D., professor. His research interests include computer vision, machine learning, pattern recognition, image inpainting, occlusion reconstruction, shape extraction, and face recognition.

GAO Yun, M.S. candidate. His research interests include computer vision and image processing.

YAN Jia-Liang, M.S. candidate. His research interests include computer vision and pattern recognition.

ZOU Bai-Han, B.S. candidate. His research interests include object detection and image segmentation.

WANG Jian-Ming, Ph.D., professor. His research interests focus on computer vision.

Background

Image defects inpainting is an important research content in the field of computer vision. It has extensive application value in the fields of virtual reality, animation production, image editing, and protection of ancient cultural relics. In recent years, as the research on deep learning technology in image processing becomes more and more mature, it can achieve good results in advanced feature extraction and image generation of the image, making the research of image defects inpainting methods based on the deep neural network becomes a key research area for research scholars. At present, deep learning-based image defects inpainting technology has produced many amazing research works, and has been applied to tasks such as 3D reconstruction, virtual reality, and movie special effects production in the field of computer vision. Meanwhile, there are still a few of challenging problems that deserve great attention, such as the inpainting quality with high-resolution images, the generalization ability of the model. To these issues, we should have a careful review to the newly developed works and advancements in the past ten years.

This paper divides the image defects inpainting methods based on the deep neural network into 5 classes. The network structure of each algorithm is analyzed and compared in detail, and the technical knowledge and innovative parts used in the network architecture are summarized. Then the loss function, advantages and disadvantages of these methods are detailed analyzed. In addition, the commonly used datasets in the field of image defects inpainting are introduced, as well as the characteristics of the image samples in each dataset. Finally, the advantages and disadvantages of each type of algorithm are compared through experimental data, and the experimental results of each type of algorithm on the corresponding data set are given.

This work is a critical component of our NSFC research project “A Research of Automatic Compensation of General Object Occlusion (No. 61771340)” hosted by LI Yue-Long. This is a project mainly devoting to solve general occlusion compensation problem. Image defects inpainting is a core task of the project which consists of automatic compensation of general object occlusion. Without image defects inpainting, we cannot solve the problem of automatic compensation of image

Image defects inpainting is a main research topic of our research group in Tiangong University. In the past few years, we have published a few papers about this topic. In the future, we will do our best to more achievements about occlusion reconstruction.